

Zum Fahrerzustand beim automatisierten Fahren:
Objektive Messung von Müdigkeit und ihre Einflussfaktoren

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Alexander Thomas Frey
aus Düsseldorf

1. Referent: Professor Dr. Mark Vollrath
2. Referentin: Professorin Dr. Meike Jipp

eingereicht am: 16.11.2020
mündliche Prüfung (Disputation) am: 20.01.2021

Druckjahr 2021

Vorveröffentlichungen

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Hohm, A., Klejnowski, L., Skibinski, S., Bengler, K., Berger, S., Vetter, J., . . . & Stürmer, T. (2018). *KO-HAF - Kooperatives Hochautomatisiertes Fahren: (Projektübergreifender Schlussbericht)*. Abgerufen von: https://www.ko-haf.de/fileadmin/user_upload/projekt/19S14002_Ko-HAF_partner%C3%BCbergreifender-Schlussbericht_final.pdf [09.04.2020].

Tagungsbeiträge

Frey, A. (2016). *Impact of Automation on Driver's Vigilance*. 4th KOTSA-BASSt Joint Transport Safety Symposium im Bundesministerium für Verkehr und digitale Infrastruktur, Berlin, 25.08.2016.

Frey, A. (2016). *Fahrervigilanz in einer monotonen, teilautomatisierten Realfahrt*. 10. Doktorandenworkshop der Fachgruppe Verkehrspsychologie der DGPs, Dortmund, 12.09 - 14.09.2016.

Frey, A. (2017). *Vigilanz beim automatisierten Fahren*. Junge Wissenschaftlerinnen und Wissenschaftler der Bundesanstalt für Straßenwesen, Bergisch Gladbach, 31.03.2017.

Frey, A. (2017). *Effekte längerer Automatisierung, Vigilanz und Müdigkeit des Menschen*. Zwischenpräsentation des BMWi-Projekts Ko-HAF, Aschaffenburg, 18.05.2017.

Frey, A. (2017). *Impact of Automation on Driver's Vigilance*. Meeting between MLIT, NTSEL and BASSt, Bergisch Gladbach, 29.08.2017.

Frey, A. (2017). *Impact of Automation on Driver's Vigilance*. Besuch der singapurischen Landesverkehrsbehörde in der BASSt, Bergisch Gladbach, 31.08.2017.

Frey, A. (2017). *Müdigkeit und Vigilanz beim automatisierten Fahren*. 12. Dortmunder AutoTag, Dortmund, 07.09.2017.

- Frey, A. (2017). *Versuchsteil I: Vigilanz in einer monotonen, teilautomatisierten Realfahrt*. 11. Doktorandenworkshop der Fachgruppe Verkehrspsychologie der DGPs, Braunschweig, 19.09. - 21.09.2017.
- Frey, A. (2018). *Müdigkeit in einer automatisierten Realfahrt*. 12. Doktorandenworkshop der Fachgruppe Verkehrspsychologie der DGPs, Aachen, 25.09. - 27.09.2018.
- Frey, A. (2019). *Müdigkeit und Vigilanz in einer automatisierten Realfahrt*. 10. VDI-Fachtagung Mensch-Maschine-Mobilität (S. 121-132). Braunschweig: DLR.
- Radlmayr, J., Feldhütter, A., Frey, A., Jarosch, O., Marberger, C., Naujoks, F., . . . & Bengler, K. (2018). Drowsiness and fatigue in conditionally automated driving—Towards an integrative framework. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting Europe Chapter, (HFES) 2018*.

Posterbeiträge

- Frey, A. (2017). *Fahrervigilanz beim Überwachen einer automatisierten Realfahrt*. Zwischenpräsentation des BMWi-Projekts Ko-HAF, Aschaffenburg, 18.05.2017.
- Frey, A. (2018). *Monitoring Automated Driving: Physiological Measurement of Humans' Fatigue*. Transportation Research Arena (TRA), Wien, 16.04. - 19.04.2018.
- Frey, A. (2018). *Monitoring Automated Driving: Measurement of Humans' Fatigue*. Abschlusspräsentation des BMWi-Projekts Ko-HAF, Rodgau-Dudenhofen, 19.09. - 20.09.2018.

Danksagung

An dieser Stelle möchte ich allen Personen danken, ohne deren Unterstützung vorliegende Arbeit so nicht möglich gewesen wäre. Mein besonderer Dank gilt Herrn Prof. Dr. Mark Vollrath für die Betreuung der Dissertation – verbunden mit spannenden Diskussionen und wegweisenden Ratschlägen zu jeder Zeit in allen Phasen. Einen großen Dank möchte ich darüber hinaus an Frau Prof. Dr. Meike Jipp für die Übernahme des Koreferats richten. Auch den Mitarbeiterinnen und Mitarbeitern der Bundesanstalt für Straßenwesen (BASt) gilt mein herzlicher Dank: Insbesondere Herr Ass. jur. Tom Michael Gasser unterstützte fachlich wie moralisch, stets ansprechbar und hilfsbereit. Herr Dipl.-Ing. Roland Schindhelm, Frau M. Sc. Elisabeth Shi und Herr M. Sc. André Wiggerich begleiteten das Vorhaben ebenfalls mit wertvollen Anregungen. Ferner sorgte Herr Dipl.-Ing. (FH) Torsten Marx mit großem Engagement für die technische Zuverlässigkeit des Versuchsfahrzeugs. Dass Videodaten valide ausgewertet werden konnten, ist den Unterweisungen von Frau Dr. Patricia Nowak (TU Braunschweig) zu verdanken. Ebenfalls möchte ich das tatkräftige Mitwirken von Frau Maike Andres, Frau M. Sc. Marie Ufert und Frau B. Sc. Michelle Wind hier hervorheben. Mein besonderer Dank gilt auch meinen Eltern, die mich immer bei allen Entscheidungen unterstützt haben und mir insbesondere mein Studium ermöglichten. Nicht zuletzt danke ich meiner Frau, die mir jederzeit den nötigen Rückhalt gab.

- Meiner Familie -

Zusammenfassung

Eine automatisierte Fahrfunktion gemäß SAE Level 2 erfordert die permanente Überwachung der Fahrzeugsteuerung und der Verkehrsumgebung, wohingegen ein Level 3 die Abwendung von der Fahraufgabe gewährleistet und somit die Beschäftigung mit fahrfremden Tätigkeiten (FFT) ermöglicht. Erkenntnisse aus Grundlagen- und Anwendungsforschung zeigen, dass der Mensch insbesondere Level 2-typische Überwachungstätigkeiten, d.h. Vigilanzaufgaben, nicht lange bewerkstelligen kann, weil er ermüdet. Dabei spielt neben schlafbezogenen Faktoren auch die sogenannte passive aufgabenbezogene Müdigkeit eine entscheidende Rolle. Letztere kann sich sowohl durch Überbeanspruchung aufgrund eines zu hohen Workloads in Verbindung mit erhöhtem Stress abzeichnen als auch durch Unterbeanspruchung in monotonen Situationen aufgrund eines zu niedrigen Arousal-Levels und Langeweile. Dieser Dualismus wurde anhand einer umfangreichen Literaturrecherche zu den Themen Vigilanz und Müdigkeit in vorliegender Arbeit aufgearbeitet und im Hinblick auf die Anforderungen an einen adäquaten Fahrerzustand beim automatisierten Fahren eingeordnet. Dabei zeichneten sich große Forschungslücken ab: Generell fehlt ein fundierter Nachweis mittels objektiver Messmethodik, dass sich der Fahrerzustand (i.S. seines Ermüdungsverhaltens) während Level 2 in einer Realfahrt kongruent zum hinlänglich bekannten Vigilanzdekrement über die Fahrdauer verhält. Speziell bleibt unbeantwortet, ob vorrangig Über- oder Unterbeanspruchung für das Vigilanzdekrement verantwortlich ist und ob eine geringe Funktionszuverlässigkeit die Vigilanz aufrechterhalten kann. Erste empirische Forschungsansätze bedienen sich lediglich subjektiven Eigenbewertungen zum Müdigkeitszustand – häufig in Fahrsimulatorstudien. Dabei werden intra- und interindividuelle Unterschiede regelmäßig als ernstzunehmende Fehlervarianzquelle moniert, jedoch bis dato nicht systematisch untersucht. An diesen offenen Fragestellungen setzt vorliegende Arbeit an.

Das primäre empirische Forschungsziel umfasste die Machbarkeit von objektiver Müdigkeitsdetektion während einer Level 2-Nutzung in einem „Wizard-of-Oz-Fahrzeug“ auf einer monotonen Teststrecke mittels Elektroenzephalografie (EEG), im Speziellen über die Analyse von EEG-Alpha-Spindeln. Als Validierungsgrundlage diente eine videobasierte Müdigkeitsbewertung. Ein erstes Experiment mit einer Nettostichprobe von $N = 35$ Probanden legte den Fokus auf die Auswirkungen der Fahrdauer (60 Minuten) in Level 2 und dessen Funktionszuverlässigkeit. Erwartungsgemäß traten große individuelle Unterschiede im Ermüdungsverhalten auf, sodass nur eine Stichprobenhälfte EEG-basiert als „ermüdet“ klassifiziert werden konnte. Diese Probanden besaßen in den ersten 20 Minuten in Level 2 eine deutliche Müdigkeitszunahme, wobei ca. 66 % der Varianz in der Ermüdungsstärke anhand der Fahrdauer aufgeklärt werden

konnten. Anschließend trat eine Müdigkeitssättigung ein. Sowohl ältere Versuchsteilnehmer als auch Fahrer unter Koffeineinfluss wiesen eine signifikant niedrigere Ermüdung auf. Nach ca. 50 Minuten zeigten „ermüdete“ Probanden verzögerte Reaktionen auf einen Automatisierungsfehler, wobei die Funktionszuverlässigkeit grundsätzlich keinen Einfluss auf das Ermüdungsmuster hatte. Allerdings konnten Automatisierungsfehler eine Aktivierung für etwa zwei Minuten im Fahrerzustand hervorrufen, wenn eine hohe Ermüdung vorlag und noch keine Habituation an die Fehler stattfand. Als relatives Müdigkeitsmaß bildete die Alpha-Spindelrate vorrangig das Ermüdungspotential der Teilnehmer ab, wohingegen videobasiert auf die absolute Müdigkeit geschlossen werden konnte.

In einem zweiten Experiment sollte evaluiert werden, ob die Beschäftigung mit motivierenden FFT in einer Level 3-Fahrt Müdigkeit reduzieren kann. Hierfür erlebten $N = 19$ Teilnehmer abwechselnd entweder ein Level 2 oder ein Level 3 kontinuierlich für jeweils 28 Minuten mit anschließenden Transitionen zum manuellen Fahren. Die Übernahmezeiten betrugen müdigkeitsbedingt nach Level 2 im Mittel ca. vier Sekunden. Wenn nach Level 3 die FFT vor Übernahme beiseitegelegt wurde, waren Übernahmezeiten von durchschnittlich ca. acht Sekunden messbar. Die Ausbildung von Müdigkeit konnte in Level 3 für bis zu 25 Minuten unterdrückt werden. Somit besitzt ein Level 3 mit FFT-Beschäftigung (verglichen mit einem Level 2) ein deutliches Aktivierungspotential, wodurch ebenfalls die Verkehrssicherheit erhöht werden kann.

Inhaltsverzeichnis

1	Einführung	12
2	Theoretischer Hintergrund.....	16
2.1	Fahrerassistenz und Fahrzeugautomatisierung	16
2.2	Vigilanz.....	21
2.2.1	Begriffs- und Konstruktdefinitionen	21
2.2.2	Reizverarbeitung	22
2.2.3	Unter- und Überbeanspruchung	24
2.2.3.1	Langeweile.....	26
2.2.3.2	Mentaler Workload und Stress	28
2.2.4	Zusammenfassung Vigilanz	30
2.3	Müdigkeit.....	31
2.3.1	Begriffs- und Konstruktdefinitionen	31
2.3.2	Entstehung von Müdigkeit	33
2.3.3	Schlafbezogene Müdigkeit.....	36
2.3.3.1	Zirkadianer Rhythmus	37
2.3.3.2	Chronotypen	38
2.3.4	Müdigkeitsmodell für SAE Level 2	39
2.3.5	Müdigkeitsindikatoren	40
2.3.5.1	Selbsteinschätzung	42
2.3.5.2	Fahrverhalten	44
2.3.5.3	Verhaltensbeobachtung	45
2.3.5.4	EEG-Alpha-Aktivität und Alpha-Spindeln	50
2.3.6	Zusammenfassung Müdigkeit	56
2.4	Fahrerzustand beim automatisierten Fahren	57
2.4.1	Überblick über Themenfelder und Forschungsbedarf.....	60
2.4.2	Unfälle während SAE Level 2	63
2.4.3	Fahrerzustand am Modell.....	64
2.4.4	Experimentelle Methoden zur Erforschung des Fahrerzustands.....	66
2.4.5	Ermüdung und Vigilanzdekrement: Studienlage	73
2.4.5.1	Studienlage für SAE Level 2	82
2.4.5.2	Studienlage für SAE Level 3	88
2.4.6	Zusammenfassung Fahrerzustand und offene Forschungsfragen	94

2.5	Zielsetzung der Arbeit	97
2.5.1	Fragestellungen Experiment 1	97
2.5.2	Fragestellungen Experiment 2	99
3	Experiment 1	101
3.1	Methode.....	101
3.1.1	Versuchsplan.....	101
3.1.2	Stichprobe	104
3.1.3	Probandenakquise	105
3.1.4	Versuchsmaterial	105
3.1.4.1	Teststrecke	105
3.1.4.2	Versuchsfahrzeug.....	106
3.1.4.3	Elektroenzephalographie (EEG)	111
3.1.4.4	Fragebögen.....	116
3.1.5	Instruktion.....	121
3.1.6	Versuchsdurchführung.....	122
3.1.7	Videobasiertes Müdigkeitsrating	127
3.1.8	Immersionsgüte.....	127
3.1.9	Inferenzstatistische Analyse	128
3.2	Ergebnisse	128
3.2.1	Deskriptive Daten	128
3.2.2	Subjektive Urteile	131
3.2.2.1	Müdigkeitseinschätzung.....	131
3.2.2.2	Vertrauen in die Automatisierungsfunktion.....	132
3.2.2.3	Workload.....	133
3.2.2.4	Stresserleben	134
3.2.2.5	Nutzungsbereitschaft und Komforterleben	136
3.2.2.6	Kinetose	137
3.2.3	EEG-Alpha-Spindelrate	138
3.2.3.1	Zeitreihen	138
3.2.3.2	Analyse über Aggregationsbereiche	147
3.2.3.3	Analyse des Einflusses von Automatisierungsfehlern	152
3.2.4	Videobasierte Müdigkeitsbeurteilung.....	156
3.3	Beantwortung der Fragestellungen.....	160

3.4	Diskussion.....	162
4	Experiment 2	182
4.1	Methode	182
4.1.1	Versuchsplan	182
4.1.2	Stichprobe	184
4.1.3	Probandenakquise	184
4.1.4	Versuchsmaterial.....	185
4.1.4.1	Versuchsfahrzeug	185
4.1.4.2	Fragebögen	188
4.1.5	Instruktion	189
4.1.6	Versuchsdurchführung	190
4.1.7	Videoanalyse	192
4.2	Ergebnisse.....	194
4.2.1	Subjektive Urteile.....	194
4.2.1.1	Müdigkeitseinschätzung	194
4.2.1.2	Nutzungsbereitschaft	195
4.2.1.3	Vertrauen in die Automatisierungsfunktion	195
4.2.1.4	Workload	196
4.2.1.5	Stresserleben.....	197
4.2.2	Objektive Metriken	197
4.2.2.1	EEG-Alpha-Spindelrate: Zeitreihen	197
4.2.2.2	Videoanalyse	200
4.2.2.3	Übernahme zum manuellen Fahren.....	204
4.3	Beantwortung der Fragestellungen	207
4.4	Diskussion.....	209
5	Fazit.....	220
6	Literaturverzeichnis	224
7	Anhang	244
Anhang A.	Fragebogen Experiment 1, Teil 1	244
Anhang B.	Fragebogen Experiment 1, Teil 2	251
Anhang C.	Fragebogen Experiment 2, Teil 1	258
Anhang D.	Fragebogen Experiment 2, Teil 2	264
Anhang E.	Instruktion für die manuelle Fahrt (Baselinefahrt) in Experiment 1 und 2 ...	276

1 Einführung

Autofahren, wie wir es kennen, wird aufhören zu existieren – viele aktuelle Forschungsarbeiten leiten mit einer revolutionären Mobilitätswende in die Thematik des automatisierten oder autonomen Fahrens ein. Zunächst ist aber festzustellen, dass die Entwicklung des Automobils immer einhergeht mit den technologischen Möglichkeiten seiner Zeit. Schon vor mehreren Jahrzehnten stellten sich Menschen eine Zukunft mit Fortbewegungsmitteln vor, die selbständig von A nach B fahren: Im populärwissenschaftlichen Magazin „Quest – Adventures in the World of Science“ wurde Ende der 80er Jahre das „Computer Car 2000“ beschrieben:

„Cars of the future will take the stress out of driving. Cars will be installed with an electronic system, which will enable them to travel at high speed, nose to bumper, without fear of collision. As soon as the car is on the guide track on the centre of the road the driver can sit back and watch an in-car video or snooze. Laser sensors will control the distance from the car in front and respond to underground indicators that replace traffic lights. A computer will ensure the vehicle follows a programmed route to the required destination.“ (Quest, 1989, zit. nach Stanton & Marsden, 1996, S. 36).

Heutzutage ist nicht nur die Vision, sondern auch die Technologie – maßgeblich geprägt von der Performanz von Computern bei stetig kleineren Baugrößen – auf einem Entwicklungsstand angelangt, der automatisiertes Fahren ermöglichen kann. Einige Funktionen, die kontinuierlich auf Bahnführungsebene operieren (vgl. Abschnitt 2.1), sind bereits bis in das mittlere Fahrzeugsegment durchgedrungen. Die Automobilindustrie und deren Zulieferer arbeiten folglich mit Hochdruck an Level 2- und Level 3-Funktionen gemäß SAE-Standard J3016 (SAE, 2018; vgl. Abschnitt 2.1), die peu à peu in neuen Fahrzeugmodellen Einzug finden sollen.

Abgesehen von dieser grundsätzlichen technologischen Weiterentwicklung besteht die Idee, das Autofahren mit zunehmender Fahrzeugautomatisierung verkehrssicherer zu gestalten. Der Mensch soll durch die Befreiung von der Fahraufgabe entlastet werden, in der Hoffnung, Unfallzahlen zu reduzieren, indem „menschliches Versagen“ als Unfallursache dezimiert wird. Die Verheißung ist groß – aber verbergen sich hinter diesen Annahmen womöglich neue sicherheitskritische Problemquellen?

SAE Level 2- und Level 3-Funktionen setzen immer noch einen Menschen als Regler für bestimmte Aufgaben in spezifischen Fahrsituationen voraus (Gasser & Westhoff, 2012). Diese Tatsache führt unweigerlich zu einer Fülle von neuartigen Forschungsfragen an die Mensch-Maschine-Interaktion, die ebenso schnell aufgeworfen werden, wie die Technologien sich weiterentwickeln. Schließlich geht mit zunehmender Fahrzeugautomatisierung eine grundlegende

Veränderung der Fahraufgabe im Vergleich zum manuellen Fahren einher: SAE Level 2-Funktionen benötigen eine fortwährende Aufmerksamkeit und Wachheit des Fahrers¹, SAE Level 3-Funktionen einen Fahrerzustand, aus dem heraus die manuelle Fahraufgabe wieder sicher wahrgenommen werden kann (Gasser et al., 2012). Demnach gewinnt die Detektion des Fahrerzustands bei Nutzung automatisierter Fahrfunktionen an Relevanz – im Schwerpunkt ist hier die Aktiviertheit und Müdigkeit des Fahrers von hohem Interesse.

Im Bereich der allgemeinspsychologischen Grundlagenforschung wurden in den letzten Jahrzehnten vielerlei Erkenntnisse zusammengetragen, die die Verhaltensweisen des Menschen beschreiben, sein Verhalten präzisieren und auch seine Leistungsgrenzen definieren. Dabei festigte sich das Bild, dass der menschliche Fahrer ein schlechter Überwacher ist (siehe z.B. Warm, Matthews & Finomore, 2018). Mit Hilfe sog. „Vigilanzaufgaben“, die das Grundprinzip einer fortwährenden Aufmerksamkeitsbindung abbilden, konnten deutliche Performanzeinbußen mit zunehmender Bearbeitungsdauer nachgewiesen werden (Parasuraman & Davies, 1976; Pigeau, Angus, O'Neill & Mack, 1995). Dieses *Vigilanzdekrement* manifestiert sich, weil der Mensch ermüdet (May & Baldwin, 2009).

Im Jahr 2018 ereigneten sich in Deutschland 2124 Verkehrsunfälle mit Personenschaden, die auf Übermüdung zurückzuführen sind (Statistisches Bundesamt (Destatis), 2020). Hierbei verunglückten 3419 Menschen, 38 Personen wurden getötet (Statistisches Bundesamt (Destatis), 2020). Tatsächlich wird der Anteil dieser Verkehrsunfälle an der Gesamtzahl, und damit die Dunkelziffer, deutlich höher sein (insgesamt ca. 15 bis 20 % aller Unfälle), da die Unfallursache „Übermüdung“ im Nachhinein nur schwierig festzustellen ist (Åkerstedt, 2000). Aus dem „E-Survey of Road users‘ Attitudes“ (ESRA2) geht hervor, dass zwischen 20 und 25 % der Befragten (in den meisten der 32 weltweit erhobenen Länder) angaben, in den letzten 30 Tagen Schwierigkeiten gehabt zu haben, während der Fahrt müdigkeitsbedingt ihre Augen geöffnet zu halten (Goldenbeld & Nikolaou, 2019). Da die amtliche Unfallstatistik bisweilen nicht zwischen Unfällen differenziert, die sich während des manuellen Fahrbetriebs und denen, die sich während Level 2 oder höher ereignet haben, ist generell unklar, welche Anteile auf die Nutzung dieser Fahrfunktionen zurückzuführen sind. In den letzten Jahren erfuhren jedoch einige Unfälle, die sich während Level 2 zutrugen, mediale Aufmerksamkeit (vgl. Abschnitt 2.4.2).

¹ Um eine vereinfachte Lesbarkeit zu gewährleisten, wird in der gesamten Arbeit nicht zwischen den Geschlechtern differenziert. Ausdrücklich sollen jedoch alle Geschlechter gleichberechtigt angesprochen sein.

Die vorliegende Arbeit widmet sich im Schwerpunkt der Erforschung des Fahrerzustands während einer Level 2-Nutzung, um herauszufinden, warum Fahrer beim Überwachen einer Level 2-Funktion ermüden. Sind Über- oder Unterbeanspruchungsmechanismen für die Leistungseinbußen verantwortlich? Können die grundlagenpsychologischen Prinzipien, deren Wirkweisen aus typischen Vigilanzaufgaben bekannt sind, auf diese anwendungsorientierte Problemstellung übertragen werden? Kann eine unzuverlässige Level 2-Funktion aufgrund auftretender Automatisierungsfehler aktivierend auf den Fahrer wirken? Bis dato gibt die Forschungsliteratur hierauf keine zufriedenstellenden Antworten, da ein inhomogenes Ursachenbild für die Leistungsreduktion gezeichnet wird: Häufig werden die Anforderungen des Fahrers bei Level 2-Nutzung nicht eindeutig von denen getrennt, wie sie bei Level 3-Nutzung bestehen. Zudem bedienen sich Forscher häufig subjektiven Müdigkeitsbewertungen, die den wahren Müdigkeitswert des Fahrers unterschätzen können. Daher bleibt auch die Forschungsfrage, nach welcher Nutzungsdauer eines Level 2 sich ein sicherheitskritisches Müdigkeitsniveau abzeichnen wird, weitestgehend ungeklärt und soll mit Hilfe einer objektiven Müdigkeitsdetektion in vorliegender Arbeit ebenfalls beantwortet werden. Hierfür wurde die Anwendung von Elektroenzephalographie (EEG) in einer automatisierten Realfahrt in Anlehnung an eine videobasierte Müdigkeitsbeurteilung validiert. Dieser Validitätsnachweis konnte bis zum jetzigen Zeitpunkt an anderer Stelle noch nicht erbracht werden.

In einem weiteren Schritt sollte der Frage nachgegangen werden, ob und wie ein vorteilhafter Fahrerzustand beim automatisierten Fahren aufrechterhalten werden kann. Zur Beantwortung dieser Forschungsfrage diene als Ausgangspunkt ebenfalls eine Level 2-Fahrt, die eine permanente Überwachung benötigt. Kann eine Befreiung von der Überwachungsnotwendigkeit bei gleichzeitiger Beschäftigung mit motivierenden „fahrfremden Tätigkeiten“ – folglich eine Level 3-Nutzung – die Ausbildung von Müdigkeit unterdrücken? Welche Unterschiede existieren im Ermüdungsverhalten zwischen einer Level 2- und einer Level 3-Fahrt? Ist die Übernahme der manuellen Fahrzeugsteuerung aus Level 2 oder Level 3 durch Müdigkeitserscheinungen beeinträchtigt oder wirken Übernahmeaufforderungen aktivierend auf den Fahrerzustand? Die Beantwortung dieser Fragestellungen ist insbesondere zur Abschätzung des Verkehrssicherheitspotentials bei (langandauernder) Nutzung solcher Funktionen von großem Wert. Zwar wurde Level 3 im Hinblick auf das Übernahmeverhalten (auch unter erhöhter Müdigkeit) in den letzten Jahren mit verschiedenen Intentionen beforscht, jedoch fehlt ein präziser Vergleich zu Level 2. Zudem fußen die meisten Befunde auf Fahrsimulatorstudien und/oder subjektiven Müdigkeitsbewertungen, statt Level 2 und 3 für Probanden in einer Realfahrt erlebbar zu machen.

Bei der Beantwortung aller Forschungsfragen wurde ein besonderes Augenmerk auf individuelle Einflussfaktoren im Ermüdungsverhalten gelegt. Obwohl in bisherigen Forschungsarbeiten immense intra- und interindividuelle Unterschiede als Fehlervarianzquelle bei der Müdigkeitsmessung zur Kenntnis genommen wurden, fand keine spezifische Analyse dieser Effekte statt. Auch diese Lücke soll auf Basis einschlägiger grundlagenorientierter Konstrukte in vorliegender Arbeit geschlossen werden.

Die folgenden Abschnitte leiten zunächst in das Themenfeld des assistierten und automatisierten Fahrens ein (Abschnitt 2.1). Anschließend wird sich der „Vigilanzaufgabe“ gewidmet, die – zumindest bei Nutzung einer Level 2-Funktion – den wesentlichen Anforderungsbereich des Fahrers definiert (Abschnitt 2.2). Bevor ein umfangreicher Überblick über Forschungsarbeiten gegeben wird, die die Erfassung des Fahrerzustands beim automatisierten Fahren zum Gegenstand hatten (Abschnitt 2.4), werden die verschiedenen Facetten des Müdigkeitskonstrukts (Abschnitt 2.3) inklusive der Indikatoren zur Quantifizierung von Müdigkeit im experimentalpsychologischen Anwendungskontext vorgestellt (Abschnitt 2.3.5). Zur Klärung der zuvor erläuterten Forschungsfragen wurden zwei empirische Realfahrtstudien auf einer Teststrecke konzipiert, um den Fahrerzustand bei Verwendung einer Level 2- und Level 3-Funktion mit Hilfe verschiedener Müdigkeitsindikatoren zu erheben. Die verwendete Forschungsmethodik (Abschnitte 3.1 und 4.1) und die ermittelten Ergebnisse (Abschnitte 3.2 und 4.2) sollten einerseits vor dem Hintergrund der fahrerseitigen Müdigkeitsentwicklung in Level 2 und 3, und andererseits in Bezug auf die Validität der Metriken zur Erfassung des Fahrerzustands diskutiert werden (Abschnitte 3.4 und 4.4).

2 Theoretischer Hintergrund

2.1 Fahrerassistenz und Fahrzeugautomatisierung

Für eine grundlegende Einordnung und Abgrenzung verschiedener Fahrerassistenz- und Automatisierungsfunktionen im Fahrzeug sollen zunächst die einschlägigen Termini definiert werden. Unter Fahrerassistenz versteht man – zumindest dem Namen nach – eine oder mehrere technische Funktionen, die den Fahrer bei der Durchführung seiner Fahraufgabe unterstützen sollen. Im EU-Forschungsprojekt „Advanced Driver Assistance Systems in Europe“ (ADASE) wurden Fahrerassistenzsysteme definiert als „...concepts to improve transport safety, efficiency and comfort without additional loads on resources (energy and land use) or on environment and quality of life.“ (Ulmer, 2004, S. 2). Mehrere Assistenzfunktionen können in einem System parallel zusammenwirken (Brookhuis, De Waard & Janssen, 2019; Bubb & Bengler, 2015). Auch die anteilige oder vollständige Übernahme der Fahrzeugsteuerung durch eine Automatisierungsfunktion soll in ihrem Kerngedanken eine Entlastung des Menschen als Fahrzeugführer schaffen, um Komfort und Sicherheit im Straßenverkehr zu erhöhen (Bundesministerium für Verkehr und digitale Infrastruktur (BMVI), 2015). Automatisierte Fahrfunktionen stellen daher nach Bubb und Bengler (2015) eine konsequente, technologische Weiterentwicklung von Fahrerassistenz dar, die ohne die vorausgegangenen Erfahrungen im Bereich der Umfeldwahrnehmung undenkbar gewesen wäre (Gasser et al., 2015).

Die Verbreitung von Fahrerassistenz- und Automatisierungsfunktionen in der deutschen Fahrzeugflotte ist fester Bestandteil der „Vision Zero“ im Handlungsfeld „Fahrzeug“ (Deutscher Verkehrssicherheitsrat (DVR), 2012), mit dem Anspruch, die Unfallzahlen zukünftig auf null zu senken. Aus der Erhebung von Gruschwitz, Hölscher, Raudszus und Schulz (2020) wird ersichtlich, welche Funktionen wie weit in die deutsche Fahrzeugflotte bis zum Jahr 2017 durchdrangen. Beispielsweise waren 48 % der Fahrzeuge mit einem Tempomat, lediglich 8 % aller Fahrzeuge mit einem adaptiven Abstandsregeltempomaten (ACC) ausgerüstet. Noch geringer fallen die Zahlen z.B. für eine Spurhaltefunktion aus (4 %). Insgesamt wird aus der Auflistung von ca. 40 verschiedenen Fahrerassistenz- und Automatisierungsfunktion deutlich, dass diese einer logischen Kategorisierung bedürfen. Häufig besteht jedoch die Problematik, dass verschiedene Automobilhersteller ihre angebotenen Funktionen unterschiedlich benennen und in einem spezifisch geschnürten „Fahrerassistenzpaket“ den Kunden anbieten. Der Klassifikationsvorschlag von Gasser, Frey, Seeck und Auerswald (2017) (vgl. Abbildung 1) kann helfen, hier dennoch einen strukturierten Überblick zu behalten.

Principle of Operation A: Informing and warning	Principle of Operation B: Continuously automating	Principle of Operation C: temporarily intervening in accident-prone situations
Only indirect influence on vehicle guidance via the driver. 1. Status information 2. Warning (abstract hazard) 3. Warning (concrete hazard)	Take direct influence on vehicle guidance (conscious activation by the driver – divided responsibilities in execution of the dynamic driving task). Always overrideable.	Preventive machine intervention in case of negative situation prediction. Either: I. driver as controller does not react conform to expectation <u>or</u> is inaccessible II. in accident-prone situations drivers/controllers cannot handle due to performance limitations

Abbildung 1: Übersicht über die Wirkweisen von Fahrerassistenz und Fahrzeugautomatisierung (aus Gasser et al., 2017, S. 3).

Es werden demnach drei basale Wirkweisen von Fahrerassistenz und Fahrzeugautomatisierung unterschieden, die die zuvor umrissenen Abgrenzungen berücksichtigen. In Wirkweise A werden alle informierenden und warnenden Funktionen einsortiert. Wirkweise B umfasst die kontinuierliche Fahrzeugautomatisierung und Wirkweise C klassifiziert Funktionen, die nur in unfallgeneigten Situationen in die Fahrzeugsteuerung eingreifen. Es bestehen zwei fundamentale Grundgedanken dieser Kategorisierung: Einerseits wird sich auf Funktionen beschränkt, die ausschließlich die Bahnführungsebene der Fahrzeugsteuerung betreffen. Nach Winner, Hakuli, Lotz und Singer (2016) wird die Fahrzeugsteuerung in die Navigationsebene, Bahnführungsebene und Stabilisierungsebene eingeteilt. Der Mensch als Fahrzeugführer wechselt dabei während einer Fahrperiode permanent zwischen diesen Ebenen. Weder eine Navigationsfunktion noch beispielsweise das prominente „ESP“ (elektronisches Stabilitätsprogramm) werden folgerichtig in dieser Kategorisierung berücksichtigt, da diese nicht der Bahnführungsebene zuzuordnen sind. Andererseits wird das zuvor beschriebene Merkmal aufgegriffen, wonach mehrere Funktionen im Fahrzeug zu einem System zusammengefasst werden können. Post hoc lässt sich das (gesamte) „Fahrerassistenzsystem“ so in einzelne Funktionen zerlegen und beschreiben: Beispielsweise kann eine Auffahrwarnung („Forward Collision Warning“) (Wirkweise A) mit einer Notbremsfunktion (Wirkweise C) in einem System kombiniert sein. Auch ein System, das nur eine einzige Funktion beinhaltet, ist vorstellbar.

Sowohl Wirkweise B als auch Wirkweise C werden fein strukturiert in einzelne „Level“ eingeteilt. Generell gilt hier: Je höher das Level, umso größer der Umfang, den die Funktion technisch leisten kann. Die einzelnen Level der Wirkweise B haben sich in den letzten Jahren im Forschungsbereich des automatisierten Fahrens zu einer festen Definitionsgröße entwickelt. Erstmals wurden diese im Rahmen der Projektgruppe „Rechtsfolgen zunehmender Fahrzeugautomatisierung“ der Bundesanstalt für Straßenwesen (BASt) in dem hierbei entstandenen Forschungsbericht von Gasser et al. (2012) benannt. Die „Society of Automotive Engineers“ (SAE) hat die Kategorisierungen aufgegriffen und entwickelt diese im Rahmen des SAE-Standards J3016 (SAE, 2018) kontinuierlich weiter. Tabelle 1 gibt einen definitorischen Überblick der Automatisierungslevel angelehnt an den SAE-Standard, deren deutsche Übersetzung an Gasser et al. (2012) orientiert ist.

Tabelle 1: Definitorischer Überblick der Automatisierungslevel angelehnt an den SAE-Standard J3016.

SAE Level	SAE Terminologie	Fahrer- und Automatisierungsaufgabe
0	No Automation	Der Fahrer führt kontinuierlich alle Fahraufgaben aus.
1	Driver Assistance	Der Fahrer führt kontinuierlich entweder die Quer- oder die Längsführung aus. Die Automatisierungsfunktion übernimmt die jeweils andere, nicht durch den Fahrer ausgeführte Quer- oder Längsführung. Der Fahrer muss die Funktion permanent überwachen, er muss jederzeit zur Übernahme der vollständigen Fahraufgabe in der Lage sein.
2	Partial Automation	Die Automatisierungsfunktion übernimmt Längs- und Querführung des Fahrzeugs für eine bestimmte Zeitspanne in einer bestimmten Domäne. Der Fahrer muss die Funktion permanent überwachen, er muss jederzeit zur Übernahme der vollständigen Fahraufgabe in der Lage sein.

3	Conditional Automation	Die Automatisierungsfunktion übernimmt Längs- und Querführung des Fahrzeugs für eine bestimmte Zeitspanne in einer bestimmten Domäne. Der Fahrer muss die Funktion nicht permanent überwachen, er muss als sog. "fallback ready user" mit ausreichender Zeitreserve zur Übernahme der vollständigen Fahraufgabe in der Lage sein.
4	High Automation	Die Automatisierungsfunktion übernimmt Längs- und Querführung des Fahrzeugs für einen bestimmte Zeitspanne in einer bestimmten Domäne. Der Fahrer muss nicht als sog. „fallback ready user“ zur Verfügung stehen.
5	Full Automation	Die Automatisierungsfunktion übernimmt die gesamte Fahraufgabe für die gesamte Fahrdauer in allen Domänen.

Die vorliegende Arbeit konzentriert sich auf den Zustand des Fahrers, während er eine kontinuierliche Automatisierungsfunktion gemäß eines SAE Level 2 bzw. SAE Level 3 verwendet. Daher soll auf diese beiden Level der Wirkweise B näher eingegangen werden. Unter einem SAE Level 2 (im Folgenden kurz Level 2) versteht man eine Automatisierungsfunktion, die sowohl die Längs- als auch die Querführung eines Fahrzeugs übernimmt. Dabei ist eine spezifische Domäne für den Funktionsbereich definiert – bspw. kann die Funktion ausschließlich auf Bundesautobahnen operieren. Dem Menschen auf dem Fahrerplatz obliegt bei Verwendung eines Level 2 die Aufgabe, die Funktion permanent zu überwachen. Falls diese fehlerhaft arbeitet, muss der Mensch jederzeit in der Lage sein, die Fahrzeugsteuerung sicher wahrzunehmen. Funktionsseitige Übernahmeaufforderungen gegenüber dem Fahrer zum manuellen Fahren erfolgen daher unmittelbar, d.h. ohne zeitlichen Vorlauf. Die Überwachung einer Level 2-Funktion verlangt vom Fahrer eine hohe Daueraufmerksamkeit (Vigilanz), wie sie typischerweise bei sog. Vigilanzaufgaben erbracht werden muss. Das Konstrukt der Vigilanz, grundlagenpsychologische Forschungsansätze sowie ihre Übertragbarkeit auf anwendungsbezogene Fragestellungen werden dafür zunächst in Abschnitt 2.2 thematisiert.

Auch eine Automatisierungsfunktion gemäß SAE Level 3 (im Folgenden kurz Level 3) kann ausschließlich in ihrer zugewiesenen Domäne, d.h. in einem konkreten Verkehrsbereich operieren. Die Funktion übernimmt dort die Längs- und Querführung des Fahrzeugs. Im Gegensatz zu einem Level 2 kann ein Level 3 das Erreichen eines Domänenendes mit zeitlichem Vorlauf

detektieren; auch ist diese Funktion in der Lage, bei technischem Versagen eine verkehrssichere Fahrt für einen Zeitbereich weniger Sekunden aufrecht zu erhalten. Dementsprechend ist dem Menschen auf dem Fahrerplatz eine andere Rolle als in einem Level 2 zugewiesen. Gemäß SAE-Definition fungiert der Nutzer eines Level 3 als „fallback ready user“, also als Rückfallebene. Der Mensch darf sich vollständig von der Fahraufgabe abwenden, eine Überwachung der Fahrt ist nicht notwendig. Ihm ist es gestattet, sich mit sogenannten „fahrfremden“ Tätigkeiten zu beschäftigen, jedoch nicht zu schlafen. In dem am 16. Juni 2017 neu geschaffenen §1b StVG (Straßenverkehrsgesetz) heißt es hierzu konkret:

„Der Fahrzeugführer darf sich während der Fahrzeugführung mittels hoch- oder vollautomatisierter Fahrfunktionen gemäß § 1a vom Verkehrsgeschehen und der Fahrzeugsteuerung abwenden; dabei muss er derart wahrnehmungsbereit bleiben, dass er seiner Pflicht nach Absatz 2 jederzeit nachkommen kann.“ (Bundesgesetzblatt Teil I Nr. 38., 2017, S. 1648)

In Absatz 2 heißt es weiter:

„Der Fahrzeugführer ist verpflichtet, die Fahrzeugsteuerung unverzüglich wieder zu übernehmen,

1. wenn das hoch- oder vollautomatisierte System ihn dazu auffordert oder
2. wenn er erkennt oder auf Grund offensichtlicher Umstände erkennen muss, dass die Voraussetzungen für eine bestimmungsgemäße Verwendung der hoch- oder vollautomatisierten Fahrfunktionen nicht mehr vorliegen.“ (Bundesgesetzblatt Teil I Nr. 38., 2017, S. 1648-1649)

Sobald die Level 3-Funktion folglich erkennt, dass eine baldige manuelle Übernahme der Fahrzeugsteuerung durch den Menschen angezeigt ist, erhält der Mensch eine Übernahmeaufforderung. Mit einem entsprechenden Zeitbudget muss der Fahrer seiner Funktion als Rückfallebene nachkommen und die Fahrzeugsteuerung übernehmen. Falls der Fahrer die Steuerung nicht rechtzeitig übernimmt, kann die Level 3-Funktion (abhängig vom Verkehrsszenario), einen „sicheren“, bzw. „risikominimalen“ Zustand im Verkehrsraum herstellen (Dierkes et al., 2019).

In vorliegender Arbeit soll insbesondere die menschliche Leistungsfähigkeit im Umgang mit einem Level 2 ausführliche Betrachtung erfahren. Diese fahrerzustandsbezogenen Anforderungen wurden bis dato nur unzureichend beforscht: Einerseits verschwimmen häufig Definitionsgrenzen zwischen Level 2 und 3, andererseits kommen nur selten objektive Messverfahren unter kontrollierten, aber dennoch realen Fahrbedingungen, zum Einsatz (siehe hierzu Abschnitt 2.4). Da der Fahrer in Level 2 insbesondere die entsprechende Wachheit besitzen muss, um die permanente Überwachungstätigkeit leisten zu können, wird im nachfolgenden Abschnitt 2.2 zunächst der Fokus auf die benötigte *Vigilanz* gelegt, bevor in Abschnitt 2.3 das Konstrukt der *Müdigkeit* grundlegend thematisiert wird.

2.2 Vigilanz

2.2.1 Begriffs- und Konstruktdefinitionen

Der Begriff der Vigilanz wurde bereits früh durch Head (1923) geprägt, der Vigilanz als generalen Zustand von zentralnervöser Vitalität (d.h. physisch wie psychisch) verstand, der differenzierte Aussagen über die Reaktionsfähigkeit auf einen Reiz erlaubt. Heute wird unter Vigilanz im Allgemeinen die Fähigkeit eines Individuums verstanden, seine Aufmerksamkeit über eine längere Zeitspanne aufrecht zu erhalten, um auf spezifische Reize (adäquat) zu reagieren (Pattyn, Neyt, Henderickx & Soetens, 2008; Scerbo, 2001; Warm, Dember & Hancock, 1996). Sogenannte Vigilanzaufgaben zeichnen sich daher dadurch aus, dass eine monotone Tätigkeit über eine lange Zeitdauer ausgeführt werden muss. Je nach Anwendungskontext muss dabei auf vorab definierte – mehr oder weniger artifizielle – Reize (sog. Zielreize), die in einem bestimmten Verhältnis zu Distraktorreizen auftreten, eine Reaktion erfolgen. Dementsprechend besitzen Vigilanzaufgaben ein spezifisches Signal-Rausch-Verhältnis.

In experimentalpsychologischen Untersuchungen können im Labor Vigilanzaufgaben so programmiert werden, dass Auswirkungen der Manipulationen valide abgebildet werden können. Die vermutlich erste Aufgabe, die mit dieser Prämisse konstruiert wurde, war der sog. „Mackworth-Clock-Test“ (Mackworth, 1948) (siehe nachfolgenden Abschnitt 2.2.2). Häufige Stellgrößen von Vigilanzaufgaben sind die Reizmodalität (Davies & Parasuraman, 1982), der Monotoniegrad der Aufgabe (Howell, Johnston & Goldstein, 1966), die Bearbeitungsdauer („time-on-task“) (Molloy & Parasuraman, 1996), das Signal-Rausch-Verhältnis (Johnston, Howell & Williges, 1969; Thomson, Besner & Smilek, 2016) und die Salienz der Zielreize relativ zu den Distraktorreizen (Hitchcock et al., 2003; Parasuraman & Mouloua, 1987). Auch die Auswirkung von (positivem oder negativem) Feedback über die korrekte Detektionsquote stellt eine beliebte Forschungsfrage dar (Becker, Warm, Dember & Hancock, 1991; Freeman, Mikulka, Scerbo & Scott, 2004). Als abhängige Variable kann das Verhältnis zwischen Treffern, korrekten Zurückweisungen, falschen Alarmen und Auslassungen fungieren, wie es die Signalentdeckungstheorie vorsieht (Craig, 1987; Green & Swets, 1966; Thomson et al., 2016). Durch Relation zwischen der Trefferrate und der Rate falscher Alarme ist es möglich, sowohl das Sensitivitätsmaß d' als auch die Antworttendenz β (bzw. das Antwortkriterium „ c “) eines Probanden zu bestimmen (Guralnick & Harvey, 1970). Häufiger werden jedoch über die Reaktionsgeschwindigkeit auf die Zielreize Rückschlüsse auf den gegenwärtigen Vigilanzzustand eines

Menschen gezogen (Körber, Cingel, Zimmermann & Bengler, 2015). Kernelement der Forschungsfragen ist es, die Entstehung von Vigilanzdekrement vorherzusagen sowie Vigilanz trotz monotoner Aufgabe über möglichst lange Zeit aufrecht zu erhalten.

Viele Jahrzehnte Vigilanzforschung wurden früh durch anwendungsbezogene Fragestellungen geprägt. Es existierten bzw. existieren diverse Arbeitsfelder, in denen Tätigkeiten, die ein hohes Maß an Vigilanz erfordern, zur täglichen Aufgabe gehören. Hier sind Überwachungsaufgaben mit hoher Sicherheitsrelevanz zu nennen, wie diese z.B. bei der Gepäckkontrolle via Röntengerät am Flughafenterminal zu leisten sind, oder Flugbeobachtungen von Fluglotsen, die Radarbildschirme auf relevante Veränderungen hin beobachten müssen. Insbesondere stellen die Aufgaben von Piloten schon einige Jahrzehnte Überwachungstätigkeiten mit hoher Verantwortung dar, daher entstammen viele Erkenntnisse zur menschlichen Vigilanz im Verkehrskontext der Luftfahrt (siehe z.B. Mouloua, Deaton & Hitt, 2001). Mit zunehmender Fahrzeugautomatisierung werden solche Aufgaben auch dem (gemeinen) Autofahrer zuteil, schließlich stellt auch das Überwachen einer automatisierten Fahrfunktion gemäß Level 2 (vgl. Abschnitt 2.1) eine Vigilanzaufgabe dar.

2.2.2 Reizverarbeitung

Das klassische Vigilanz-Paradigma, das in dieser Form bereits von Mackworth (1948) angewandt wurde, sah folgende Merkmale vor: Zielreize mit einer geringen Salienz mussten in einem monotonen Versuchsaufbau über eine lange Zeitdauer identifiziert werden. Beim prominenten „Mackworth-Clock-Test“ (Mackworth, 1948) wird der Proband instruiert, die Bewegung des Sekundenzeigers einer Uhr zu überwachen. Immer dann, wenn der Sekundenzeiger zwei statt einer Sekunde weiterspringt, muss der Proband eine Reaktion zeigen (einen Taster drücken). Mackworth (1948), aber auch andere Autoren, die diese oder eine ähnlich gestaltete Aufgabe verwendeten, berichten relativ übereinstimmend, dass Vigilanzdekrement ein kontinuierlicher Prozess ist, und bereits nach ca. 15 Minuten Bearbeitungsdauer deutlich (im Feldversuch in abgeschwächter Form) in Erscheinung tritt (Parasuraman & Davies, 1976; Pigeau et al., 1995). Dieses manifestiert sich in einer Erhöhung der Reaktionszeit, Anzahl der Auslassungen sowie der falschen Alarme. Bei diesem Aufgabentyp liegt ein „top-down“-Prozess der Aufmerksamkeitskontrolle vor: Der Proband entscheidet (wurde natürlich so instruiert), auf welchen der präsentierten Reize eine entsprechende Reaktion erfolgen soll. Die Aufmerksamkeitskontrolle erfolgt hierbei endogen – das Ausrichten der Aufmerksamkeit steht unter Kontrolle des Probanden. Kontrollierte, endogene Reizverarbeitung erfordert eine hohe kognitive Leistung des Probanden, begrenzt durch dessen Arbeitsgedächtniskapazität, und läuft seriell ab

(Schneider & Shiffrin, 1977). Diese selektive, visuelle Aufmerksamkeitsleistung ist nach Posner (1980) auf einen definierten Bereich bestimmter Größe beschränkt. Das Lichtkegelmodell veranschaulicht, dass dieser von Aufmerksamkeit „ausgeleuchtete“ Bereich unter Kontrolle des Probanden verschoben werden kann (Posner, 1980). Auf diese Weise – Broadbent (1958) würde hier von einer frühen Filterung sprechen – kann mit Hilfe von Hinweisreizen („cueing“) die Aufmerksamkeit frühzeitig auf die Zielreize gelenkt werden. Dieser „bottom-up“-Prozess der Aufmerksamkeitsausrichtung obliegt nicht mehr der vollständigen Kontrolle des Probanden, sondern läuft weitgehend automatisiert ab. Nach Schneider und Shiffrin (1977) benötigt dieser exogene Prozess, der parallel ausgeführt werden kann, nur eine geringe kognitive Leistung, da das Arbeitsgedächtnis so gut wie gar nicht beansprucht wird.

Ein Beispiel einer weitverbreiteten Vigilanzaufgabe, die mit automatisierter (exogener) Aufmerksamkeitsausrichtung arbeitet, ist die „psychomotor vigilance task“ (PVT; Dinges & Powell, 1985). Die PVT ist ein beliebtes Werkzeug, um Veränderungen (vor allem Einbußen) aufgrund von Schläfrigkeit (häufig im Zuge von Schlafdeprivation) in der kognitiven Performanz messbar zu machen (Dorrian, Rogers & Dinges, 2004). Auch Auswirkungen von Gegenmaßnahmen wie sehr kurze Schlafepisoden („naps“) (Dinges, Orne, Whitehouse & Orne, 1987) oder Koffeineinnahme (Wright, Badia, Myers & Plenzler, 1997) auf die Schläfrigkeit werden gerne mit der PVT bestimmt. Während der 5- oder 10-minütigen Aufgabe (Lamond, Dawson & Roach, 2005) ist eine schnellstmögliche Reaktion (Tastendruck) auf einen kleinen, roten Lichtreiz erforderlich, der immer an einem konstanten Ort erscheint. Seine Reaktionsgeschwindigkeit wird dem Probanden unmittelbar zurückgemeldet. Die Instruktion sieht vor, die Reaktionsgeschwindigkeit so niedrig wie möglich zu halten, jedoch nicht zu früh zu reagieren; das Interstimulusintervall variiert typischerweise zwischen zwei und zehn Sekunden (Dorrian et al., 2004). Kongruent zu endogenen Vigilanzaufgaben, werden sowohl Reaktionszeiten als auch Auslassungen und falsche Alarmer analysiert. Allerdings ist der Leistungsverlauf über die Bearbeitungsdauer von geringerem Interesse – vielmehr gilt es, die Leistung des Probanden zu verschiedenen Messzeitpunkten zu beurteilen (Drummond et al., 2005). Dorrian et al. (2004) listen ausführlich auf, dass sich verschiedene Beeinflussungen schlafbezogener Müdigkeit in degradierender PVT-Performanz niederschlagen. Hier sind vor allem Schlafdeprivation, Zeit-zonenwechsel (Jetlag), Schlafapnoe, Beleuchtung und Koffeinzufuhr zu nennen.

Gunn et al. (2005) haben im ingenieurpsychologischen Anwendungskontext beide Formen, endogene und exogene Aufmerksamkeit, in einer Vigilanzaufgabe vergleichend untersucht. Die Probanden sollten den Flug unbemannter Flugobjekte (Drohnen) überwachen und kritische Ge-

fahrensignale detektieren. Hierfür standen ihnen zwei verschiedene Displayvarianten zur Verfügung, auf denen Zahlenpaare eingeblendet wurden. In der endogenen Bedingung (hier kognitiv genannt) galt es für den Probanden nur dann zu reagieren, wenn eine der beiden Zahlen gerade, die andere jedoch ungerade war. Die exogene (hier physische) Displayvariante zeichnete sich hingegen dadurch aus, dass Zielreize aus unterschiedlich groß dimensionierten Zahlen bestanden, die Zahlenwerte selbst jedoch keine Rolle spielten. Gunn et al. (2005) berichten, dass – unabhängig von der Displayvariante – ein signifikantes Vigilanzdekrement über die Zeit (ca. 32 Minuten) zu beobachten war, das jedoch unter Verwendung des physischen „Vigilanz-Displays“ signifikant niedriger war als beim kognitiven „Vigilanz-Display“. Neben mehr Treffern und weniger falschen Alarmen konnte in der exogenen Aufmerksamkeitsbedingung auch eine geringere kognitive Belastung gemessen werden, weswegen solche Anzeigen im Anwendungsfall zu bevorzugen seien. Diese Ergebnisse bekräftigen einerseits die Annahme, dass kognitive Prozesse, die bei Vigilanzaufgaben wirken, sowohl einen endogenen als auch exogenen Anteil besitzen (Jacoby, 1991). Andererseits wird auch erkennbar, dass das Vigilanzdekrement bei Überwachungstätigkeiten über die Zeit einen vornehmlich endogenen aufmerksamkeitsbezogenen Ursprung hat. Dies stützt die Annahmen, die bereits Fisk und Schneider (1981) zum primären Mechanismus des Vigilanzdekrements machten, wonach *kontrollierte* Reizverarbeitung vornehmlich für die Leistungsreduktion verantwortlich ist.

2.2.3 Unter- und Überbeanspruchung

Selbst wenn sich der zugrundeliegende Aufmerksamkeitsprozess für Vigilanzdekrement damit klarer abzeichnet, bleibt weiterhin die Frage offen, warum genau eine Aufmerksamkeitsreduktion eintritt. Pattyn et al. (2008) stellen zwei kognitionspsychologische Theorien vor, die die auftretenden Aufmerksamkeitseinbußen ergründen möchten. Einerseits besteht die Vermutung, dass Vigilanzdekrement die Folge eines zu geringen Erregungsniveaus (bzw. Reizverarbeitungsniveaus) im Überwachungssystem ist, das durch die monotonen Eigenschaften typischer Vigilanzaufgaben hervorgerufen wird. Andererseits wird hypothetisiert, dass die Leistungsreduktion aus einer Abnahme der Aufmerksamkeitskapazität über die Zeit resultiert, da die mentale Beanspruchung (der „mentale Workload“) bei Vigilanzaufgaben zu hoch ist. Aus diesem Dualismus heraus kann sowohl eine „Unterbeanspruchungshypothese“ als auch eine „Überbeanspruchungshypothese“ formuliert werden (Pattyn et al., 2008). Befürworter der Unterbeanspruchung („mindlessness theory“; Körber et al., 2015) erklären, dass Monotonie in der Aufgabe und die empfundene Langeweile (vgl. Abschnitt 2.2.3.1) ein kognitives Abschweifen von der Aufgabe bewirkt (Hill & Perkins, 1985; Pattyn et al., 2008). Dagegen verweisen die Unterstüt-

zer der Überbeanspruchung („resource theory“, nach Körber et al., 2015) darauf, dass Aufmerksamkeitsressourcen bei Vigilanzaufgaben im hohen Maße beansprucht werden (vgl. Abschnitt 2.2.3.2), da die Diskrimination zwischen Ziel- und Distraktorreizen so schwierig ist. Dadurch werden Vigilanzaufgaben als sehr anstrengend erlebt (Grier et al., 2003; Helton & Russell, 2012; Warm et al., 1996; Warm, Parasuraman & Matthews, 2008).

Pattyn et al. (2008) bezeichnen den Dualismus zwischen Unter- und Überbeanspruchungshypothese als ein Paradoxon. Dem soll an dieser Stelle widersprochen werden. Tatsächlich können beide Theorien parallel existieren, wie die Modellannahmen von Fairclough (2001) zeigen. Demnach kann die Bearbeitung von Vigilanzaufgaben, aufgrund ihrer zuvor beschriebenen Merkmale, eine Unterbeanspruchung bewirken. Der Bearbeiter muss daraufhin seine kognitive Leistung erhöhen, um eine adäquate Performanz zu erreichen – sofern er natürlich hierzu instruiert bzw. motiviert ist. Fairclough (2001) erklärt, dass der Bearbeiter eine Balance zwischen adäquater Performanz und Komfort erzielen möchte. Wenn in einer Vigilanzaufgabe bei längerer Bearbeitungsdauer („time-on-task“) ein hohes Leistungsniveau erreicht werden soll, ist unweigerlich eine höhere mentale Anstrengung („mental effort“) notwendig. Dieser manifestiert sich in einer messbaren Überbeanspruchung (höherer Workload) des Bearbeiters. Ebenfalls ist dieses Regulationsprinzip mit dem etwas älteren Modell von Hockey (1997) vereinbar. Demnach findet ein routinemäßiger, introspektiver Abgleich der gezeigten Performanz in einer (Vigilanz-) Aufgabe mit dem anvisierten Ziel statt (Schleife A). In Schleife B wird der Workload überprüft, der benötigt wird, um die geforderte Performanz zu erbringen. Als Konsequenz muss im Regelkreis entweder das Aufgabenziel angepasst, oder aber der Workload erhöht werden, um auch bei Unterbeanspruchung ein moderates Performanzlevel zu gewährleisten (Hockey, 1997).

Auf den Fahrkontext übertragen sollte die Fahreraufgabe demnach in ihrer Beanspruchung des Fahrers ausgewogen konzipiert sein. Auf Vigilanzaufgaben trifft dies jedoch aufgrund ihrer zuvor thematisierten Charakteristik mitnichten zu. Stattdessen wird der Fahrer in die Randbereiche der Beanspruchung gezwungen. Die in der Theorie notwendige Ausbalancierung zwischen Unter- und Überbeanspruchung kann mit dem viel zitierten „Yerkes-Dodson-Gesetz“ (Yerkes & Dodson, 1908) veranschaulicht werden (vgl. Abbildung 2).

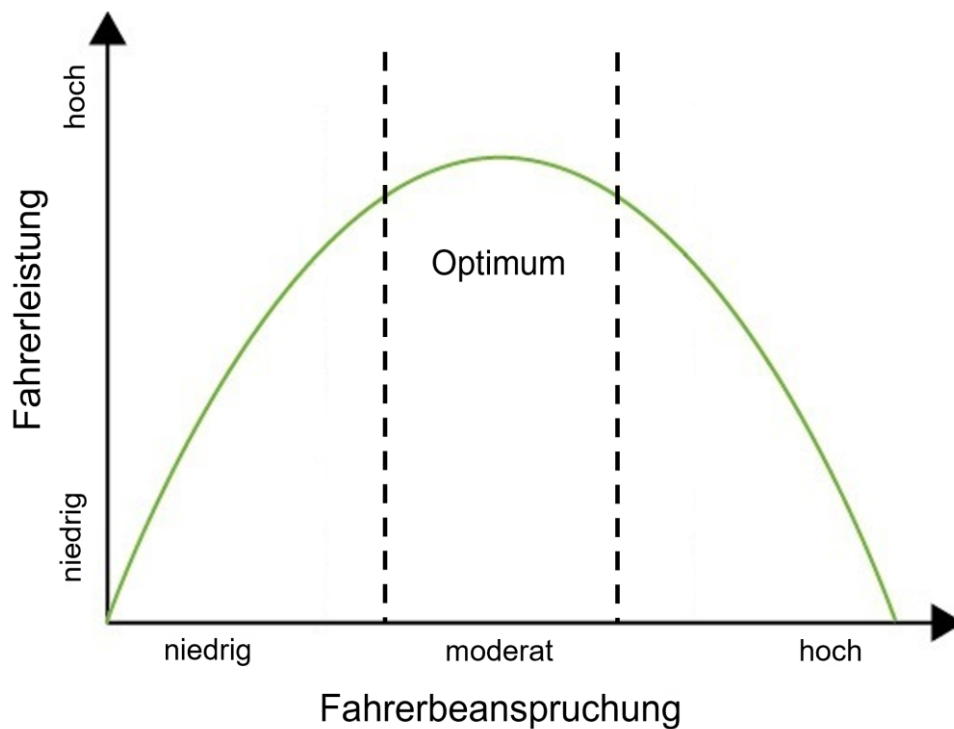


Abbildung 2: Yerkes-Dodson-Gesetz in Anlehnung an Teigen (1994).

In seiner ursprünglichen Form wurde hiermit das Lernverhalten von Mäusen beschrieben, wonach die Lernleistung bei moderater Elektroschockintensität am höchsten war, jedoch bei sehr niedriger oder sehr hoher Stimulation gering ausfiel (Yerkes & Dodson, 1908). Teigen (1994) erklärt, dass dieses Gesetz aber auf vielerlei psychologische Prinzipien angewandt werden kann. Konkret auf Vigilanzaufgaben übertragen, wäre einerseits dafür zu sorgen, dass die Merkmale der Aufgabe keine zu geringe, andererseits aber auch keine zu hohe Beanspruchung hervorrufen – ansonsten besteht die Gefahr, dass der Fahrer ermüdet. Diese Problematik wird in Abschnitt 2.3.2 behandelt. Zunächst soll sich der Interaktion zwischen Unter- und Überbeanspruchung bei Vigilanzaufgaben über die Konstrukte „Langeweile“, sowie „Mentaler Workload und Stress“ genähert werden.

2.2.3.1 Langeweile

Geiwitz (1966) definiert Langeweile als Gefühl verminderter Erregung, erhöhten Zwangs, hoher Wiederholungshäufigkeit und Unzufriedenheit. Auch Merrifield (2010) unterstreicht diese vier Facetten, ordnet diese einem Zustand der Apathie („...disconnection from the world around...“; Merrifield, 2010, S. 4) (siehe auch Cunningham, Scerbo & Freeman, 2000) und der Unruhe („...desires to engage in meaningful activity, but is unable to do so...“; Merrifield, 2010, S. 4) zu. Dies ist auch weitestgehend mit dem Langeweile-Modell von Hill und Perkins

(1985) vereinbar, wonach Langeweile sowohl eine kognitive Komponente (die subjektive Monotonie) als auch eine affektive Komponente (ein hohes Frustrationslevel) besitzt. Eine notwendige Bedingung für das Erleben von Langeweile ist demnach, dass eine Situation über monotone Eigenschaften verfügen muss (Scerbo, 2001). Vigilanzaufgaben erfüllen dieses Merkmal. Die Forschungsergebnisse von Scerbo (1998) zeigen, dass Probanden das maximale Level von Langeweile während einer Vigilanzaufgabe nach zehn bis 15 Minuten erreichen. Übliches Kriterium zur Bestimmung des maximalen Langeweile-Niveaus ist der Abbruchzeitpunkt der Aufgabe (Scerbo, 2001) oder iterative Befragungen über die gegenwärtig empfundene Langeweile (Gordon, Wilkinson, McGown & Jovanoska, 1997). Betrachtet man das übliche Performanzdekrement bei Vigilanzaufgaben, das häufig bereits nach 15 Minuten Bearbeitungsdauer festzustellen ist (Parasuraman & Davies, 1976), so wird deutlich, dass sich die Unterbeanspruchungsfacette dieser Aufgaben in Form von Langeweile widerspiegeln kann. Braby, Harris und Muir (1993) stellen diesen Zusammenhang ebenfalls nicht in Frage, relativieren jedoch, dass Langeweile kein Synonym für Unterbeanspruchung darstellt. Hier muss zwangsläufig zwischen niedrigem Arousal-Level in kognitiver Hinsicht und dem Gefühl der Langeweile differenziert werden. Erschwerend kommt hier hinzu, dass eine hohe interindividuelle Varianz (bzw. Toleranzschwelle) zwischen Menschen in ihrer Empfindung von Langeweile besteht (Farmer & Sundberg, 1986).

Pattyn et al. (2008) behaupten hingegen, dass Vigilanzaufgaben ausschließlich Langeweile messen (dieser Ansicht wurde sich bereits zuvor entgegengestellt), und verweisen auf aufgabenfremde Kognitionen („task-irrelevant / unrelated thoughts“; TUT), die sog. „Mind-wandering“ bzw. „daydreaming“ (Tagträumen) bedingen. Tatsächlich wird häufig von „Mind-wandering“ bei der Bearbeitung von Vigilanzaufgaben berichtet (Antrobus, Singer, Goldstein & Fortgang, 1970; Manly, Robertson, Galloway & Hawkins, 1999; Ralph, Onderwater, Thomson & Smilek, 2016), auch beim monotonen manuellen Fahren (Yanko & Spalek, 2014).

Cunningham, Scerbo und Freeman (2000) setzten sich das Ziel, das Auftreten von Tagträumen bei Vigilanzaufgaben elektrophysiologisch nachzuweisen. Hierzu bearbeiteten Probanden eine visuell-auditive monotone Diskriminationsaufgabe, bei denen sowohl die längere zweier vertikaler Linien als auch der lautere zweier Töne identifiziert werden sollten. Das Experiment bestand aus vier 10-minütigen Perioden, in denen die Probanden nicht nur ihre bestmögliche Leistung zeigen sollten, sondern auch das Auftreten von Tagträumen (durch Betätigung einer Taste) markieren sollten. Im Nachhinein konnte auf diese Weise ein Fenster von 30 Sekunden Länge vor und nach jeder Tagtraumepisode bestimmt werden. Innerhalb dieses Fensters wurden Daten des Elektroenzephalogramms (EEG) mit verschiedenen Frequenzbandverhältnissen analysiert

(eine detailliertere Diskussion dieser Ergebnisse findet sich in Abschnitt 2.3.5.4). Zunächst replizierten Cunningham et al. (2000) das prominente Performanzdekrement, wobei ausschließlich zwischen den ersten beiden 10-minütigen Perioden eine Verringerung der Performanz zu beobachten war. In den nachfolgenden Perioden blieben sowohl die Reaktionszeit als auch die Trefferrate, die Sensitivität und das Antwortkriterium auf einem vergleichbaren Niveau (niedriger als in der ersten Periode). Hinsichtlich der Anzahl der Tagträume war ein der Performanz kongruentes Muster zu beobachten. Demnach vermerkten die Probanden in den letzten drei Perioden signifikant mehr Tagträume als in der ersten Periode. Die Ergebnisse aus den EEG-Daten lassen weitere Ableitungen zu: Entsprechend der Hypothese konnte eine signifikant höhere Aktivierung nach einem Tagtraum als vor einem Tagtraum gemessen werden, da der Tagtraum das Resultat niedriger Aktiviertheit sei, jedoch das Bewusstwerden des Traums wieder eine höhere Aktiviertheit hervorruft. Damit können Cunningham et al. (2000) mit ihren Ergebnissen die Behauptung von Antrobus et al. (1970) stützen, wonach „Mind-wandering“ eine Art „Schutz-Funktion“ bei Langeweile darstellt, und stattfindet, um die Aktiviertheit während einer Unterbeanspruchung möglichst hoch zu halten.

2.2.3.2 Mentaler Workload und Stress

Für die Definition von mentalem Workload behelfen sich Wickens, Hollands, Banbury und Parasuraman (2013) mit einem Vergleich in Bezug auf physische Beanspruchung: „Mental workload characterizes the demands of tasks imposed on the limited information processing capacity of the brain in much the same way that physical workload characterizes the energy demands upon the muscles.“ (Wickens et al., 2013, S. 347). Dies impliziert, dass das Individuum – im hiesigen Anwendungsfall der Fahrer – eine spezifische mentale Leistungsfähigkeit besitzt. Die zu erfüllende (Vigilanz-) Aufgabe besitzt ihrerseits eine spezifische Schwierigkeit, bzw. belastet aufgrund ihrer Charakteristik die mentalen Ressourcen des Fahrers in bestimmtem Maße. Die Performanz in der Aufgabe spiegelt letztlich die Relation zwischen Aufgabenschwierigkeit und Fahrerfähigkeit wider (Moray, 1967). Warm et al. (2008) fassen kompakt zusammen, dass mentaler Workload als Beanspruchungsgrad der Informationsverarbeitungskapazität verstanden werden kann. Auch Überbeanspruchung kann somit auftreten, die ihrerseits zu Ermüdung führen kann (Hargutt, 2007; May & Baldwin, 2009; vgl. Abschnitte 2.3.2 und 2.3.4).

Das mit Abstand am häufigsten verwendete Testinstrument zur Bestimmung von Workload im Allgemeinen ist der „NASA-Task Load Index“ (NASA-TLX; Hart & Staveland, 1988). Mit

diesem Werkzeug ist es möglich, zwischen Workload, den der Bearbeiter der Aufgabe empfindet (mental, physisch und zeitkritisch) und Workload, der aus der Interaktion zwischen Bearbeiter und Aufgabe resultiert (Leistung, Anstrengung und Frustration), zu differenzieren (Warm et al., 2008). In Anwendung auf Vigilanzaufgaben konnten Warm et al. (1996) beobachten, dass das typische Vigilanzdekrement mit linearem Anstieg des subjektiven (mittels NASA-TLX gemessenen) Workloads einhergeht. Der benötigte Workload ist dabei umso höher, je geringer die Reizsalienz ist, je unsicherer der Beobachter hinsichtlich der räumlichen Lokalisation des Zielreizes ist und je mehr Distraktorreize relativ zu den Zielreizen auftreten (Warm et al., 2018). Besonders hohe Ausprägungen auf den Subskalen „mentale Beanspruchung“ und „Frustration“ des NASA-TLX kennzeichnen den benötigten Workload bei Vigilanzaufgaben (Warm et al., 1996).

In Einklang mit der multiplen Ressourcen-Theorie von Wickens (2008) zeigten Caggiano und Parasuraman (2004), dass es zu Performanzinterferenzen zwischen einer Vigilanzaufgabe und einer Arbeitsgedächtnisaufgabe kommen kann: Sobald Vigilanz- und Arbeitsgedächtnisaufgabe in derselben Modalität ausgeführt wurden (hier visuell-räumlich), verschlechterte sich die Leistung in der Vigilanzaufgabe über eine Zeitspanne von 20 Minuten signifikant (Vigilanzdekrement). Hingegen wurde keine signifikante Performanzreduktion in der visuell-räumlichen Vigilanzaufgabe gemessen, wenn eine rein visuelle Arbeitsgedächtnisaufgabe parallel bearbeitet werden musste. Dieses Ergebnis unterstreicht, dass ein hoher mentaler Workload bei der Bearbeitung von Vigilanzaufgaben vorherrscht, allerdings ressourcenspezifisch ist. Warm et al. (1996) versuchen das Paradoxon zwischen Unter- und Überbeanspruchung mit Hilfe des „Arousal-Modells“ von Vigilanz aufzulösen: Einerseits bedingt die monotone Aufgabe, dass die ausbleibende Stimulation des zentralen Nervensystems zu einer Reduktion von Aufmerksamkeit und Wachheit führt – das Vigilanzdekrement tritt ein. Andererseits wird daraufhin ein hohes Maß an Anstrengung (Ressourcen) zur Kompensation der Unterbeanspruchung benötigt, also ein hoher mentaler Workload, der die Überbeanspruchung widerspiegelt. Hier existiert eine große Schnittmenge mit den populären „Ironies of Automation“ (Bainbridge, 1983): „Although automation is designed to reduce the workload of operators, it may place them at a functional disadvantage through understimulation.“ (Warm et al., 1996, S. 185).

Der von Probanden bei Vigilanzaufgaben eingeschätzte Workload stellt eine subjektive Beurteilung dar. Das bedeutet, es können verschiedene moderierende Faktoren in Erscheinung treten, die die Bewertung des benötigten Workloads beeinflussen. Nach Scerbo (2001) und Warm et al. (2018) muss insbesondere das Konstrukt des Stresses hierbei berücksichtigt werden. Ein wesentlicher Anhaltspunkt verbirgt sich in der zuvor thematisierten Beobachtung, dass bei der

Bearbeitung von Vigilanzaufgaben Frustration erlebt wird: Aus dem Blickwinkel der Sozialpsychologie ist Stress definiert als „Die negativen Gefühle und Überzeugungen, die immer dann entstehen, wenn man sich außerstande sieht, die Anforderungen seiner Umwelt zu bewältigen.“ (Aronson, Wilson & Akert, 2008, S. 559). In der Arbeitsgruppe um Richard Lazarus (Folkman, Lazarus, Dunkel-Schetter, DeLongis & Gruen, 1986) wurde das Konstrukt des Stresses bereits differenzierter ausgearbeitet: In einer primären Bewertung wird die Relevanz eines Ereignisses, einer Situation oder einer Aufgabe (d.h. der Stressor) von der Person analysiert. Falls kritische Relevanz besteht, werden in einer sekundären Bewertung verfügbare Ressourcen für den Umgang mit dem Stressor abgerufen. Wenn keine ausreichenden Ressourcen vorliegen, entsteht Stress. Auf eine Vigilanzaufgabe – oder konkret auf das Überwachen einer automatisierten Fahrfunktion übertragen – heißt das, dass sobald die (mentalen) Ressourcen (gemessen über den benötigten Workload) erschöpft sind, und Überbeanspruchung vorliegt, ist von Stressempfinden des Fahrers auszugehen. Folkman et al. (1986) erklären, dass die Person in der Konsequenz ein bestimmtes Verhalten im Umgang mit dem Stress zeigt – die sog. „Coping-Strategie“. Dies kann entweder problemorientiert geschehen, indem die Interaktion mit dem Stressor verändert wird, oder bewertungs- bzw. emotionsorientiert, indem die Einstellung gegenüber dem Stressor eine Anpassung erfährt. Bei Bearbeitung einer monotonen Vigilanzaufgabe müssen, wie zuvor thematisiert, viele Ressourcen für die Aufrechterhaltung der Aktiviertheit oder auch Wachheit aufgebracht werden. Besteht das problemorientierte „Coping“ des Fahrers darin, die Aufgabe nicht mehr bestmöglich auszuführen (Vermeidung), wird dies zwangsläufig eine Reduktion der Aktiviertheit aber einen zusätzlichen Anstieg von Müdigkeit zur Folge haben, jedoch den benötigten Workload entlasten (Matthews & Desmond, 2002; vgl. Abschnitt 2.3.2). Demnach steigt Stress proportional mit dem aufzuwendenden Workload bei Vigilanzaufgaben über die „time-on-task“ an. Ein moderierender Faktor auf Persönlichkeitsebene ist dabei Neurotizismus; höhere Ausprägungen auf diesem Trait gehen mit höherer Wahrscheinlichkeit für Stressempfänglichkeit und höherer Müdigkeit im Fahrkontext einher (Matthews & Desmond, 1998; Matthews et al. 1999).

2.2.4 Zusammenfassung Vigilanz

Eine Automatisierungsfunktion gemäß Level 2 übernimmt zwar die Längs- und Querverführung des Fahrzeugs, allerdings muss der Fahrer zur vollständigen Übernahme der Fahraufgabe jederzeit bereit sein. Damit obliegt dem Fahrer eine permanente Überwachung der Fahrzeugsteuerung und der Verkehrsumgebung. Da bei hoher Funktionszuverlässigkeit nur seltene Interventionen des Fahrers notwendig sind, muss er folglich eine Vigilanzaufgabe bewerkstelligen: Vigilanzaufgaben erfordern eine kontinuierliche Aufrechterhaltung von Aufmerksamkeit und sind

monoton. Auf seltene Reize muss eine adäquate Reaktion erfolgen, häufig über eine lange Zeitdauer hinweg. Anhand der kognitiven Prozesse, die hierbei wirken, wird deutlich, dass vorwiegend kontrollierte, endogene Reizverarbeitung für eine Leistungsreduktion (das Vigilanzdekrement) über die Zeit verantwortlich ist. Als Ursache für die Aufmerksamkeitsreduktion kann sowohl ein zu geringes Erregungsniveau als auch eine zu hohe mentale Belastung angesehen werden (Unter- und Überbeanspruchung). Für den Bearbeiter einer Vigilanzaufgabe macht sich dies durch die Herausbildung von Langeweile aber auch Stress bemerkbar. Diese Konstrukte können einerseits durch das Auftreten aufgabenfremder Kognitionen und andererseits durch Messung mentaler Beanspruchung sichtbar gemacht werden. Im nachfolgenden Abschnitt wird die *Müdigkeit* des Fahrers thematisiert, um den Ursachen von Vigilanzdekrement während der Level 2-Nutzung auf die Spur zu kommen.

2.3 Müdigkeit

2.3.1 Begriffs- und Konstruktdefinitionen

Der Begriff der Müdigkeit ist Alltagssprachlich weit verbreitet – schließlich erfährt jeder Mensch tausendfach in seinem Leben durch Arbeit erschöpft zu sein, sich so sehr zu langweilen, dass man gähnt oder einfach nur das immer wiederkehrende Bedürfnis nach Schlaf zu empfinden. In jeder dieser Situationen ist man müde, oder man wird müde; in jedem Fall ist der Mensch in der Lage, diesen Zustand in Worte zu fassen. Das persönliche Wissen, dass dieser energetische Zustand existiert, ist für jeden Menschen unbestritten und wird als reale Entität angesehen (Polanyi, 1958, zit. nach Desmond & Hancock, 2001). Dies in Messdaten zu erfassen, ist jedoch eine – nicht selten als unlösbar bezeichnete – Herausforderung der Psychologie (Muscio, 1921; Desmond & Hancock, 2001). Unklar ist nämlich, was eigentlich gemessen werden soll. An dieser mangelhaften Quantifizierbarkeit scheitert nach Desmond und Hancock (2001) eine saubere Begriffsdefinition, da üblicherweise ein konkretes (menschliches) Verhalten oder ein Konstrukt mit Hilfe der jeweiligen Messgrößen beschrieben wird: „...this implies, tautologically, that we have to define what it is we propose to define.“ (Desmond & Hancock, 2001, S. 458). Die Konsequenz sind komplexe Be- und Umschreibungen von Müdigkeit in der Forschungsliteratur, die unter großen Anstrengungen vorgenommen werden. An dieser Stelle soll das Ausmaß dieser Definitionsversuche verdeutlicht werden.

Im englischen Sprachgebrauch wird der Begriff „Müdigkeit“ meist mit „fatigue“, hingegen „Schläfrigkeit“ mit „sleepiness“, „drowsiness“ oder „somnolence“ übersetzt. Dinges (1995) be-

hauptet, dass mit dem Terminus „sleepiness“ die präziseste Beschreibung des Konstrukts ermöglicht wird: Die Umschreibung des Bedürfnisses nach Schlaf, das durch den zugrundeliegenden neurobiologischen Prozess, der den Zirkadianen Rhythmus taktet, hervorgerufen wird. „Fatigue“, so Dinges (1995), kann hingegen mehrere Bedeutungen annehmen. Beispielsweise spricht man im Bereich der Ingenieurwissenschaften von „Materialermüdung“. Auch Johns (1998) erkennt den neurobiologischen Prozess – allerdings verwendet er den Terminus „drowsiness“, um einen Übergangszustand zwischen Wachheit und Schlaf zu definieren. Lal und Craig (2001) schließen sich der Beschreibung dieses Übergangszustands an, wobei sie definitorisch „drowsiness“ durch „fatigue“ ersetzen. Schmidt (2010) besetzt später den Begriff „drowsiness“ als *Konsequenz* erhöhter Müdigkeit. Mit der Annahme, dass „fatigue“ lediglich den Übergang zwischen Wachheit und *Schläfrigkeit* definiert, führen Desmond und Hancock (2001) eine noch feiner gegliederte Bezeichnung ins Feld. Dinges (1995) gibt in seiner Argumentation schließlich zu bedenken, dass die weite Verbreitung des Terminus „fatigue“ im Sprachgebrauch für dessen Verwendung spricht und arbeitet mit „fatigue“ und „sleepiness“ letztlich synonym. Dieser Argumentation folgt Reinprecht (2013) nicht, stattdessen gibt er eine klare Trennung zwischen Müdigkeit, die der auszuführenden Aufgabe geschuldet ist („aufgabenbezogene Müdigkeit“ nach May und Baldwin (2009)), d.h. „fatigue“, und Schläfrigkeit, die dem natürlichen Bedürfnis nach Schlaf geschuldet ist, d.h. „sleepiness“, vor. Es ist also festzuhalten, dass es nicht den einen Terminus für den Zustand der „Müdigkeit“ gibt und dass zwischen verschiedenen Autoren kein Konsens über die Verwendung der Begrifflichkeiten existiert (Weinbeer et al., 2017). Gängige Praxis ist es häufig, der „Müdigkeit“ in jeder wissenschaftlichen Arbeit von Neuem eine vertretbare und der Forschungsfrage angemessene Definition zu verleihen. Dem steht Phillips (2015) kritisch gegenüber, da im experimentellen Kontext häufig eine Gleichsetzung bestimmter Messgrößen mit Müdigkeit per se stattfindet, obwohl die Erfassung dieser Indikatoren womöglich nur Konsequenzen erhöhter Müdigkeit abbildet. Nach ausgiebiger Literaturanalyse schlägt Phillips (2015) eine komplexe Müdigkeitsdefinition vor, die dem komplexen Charakter des Konstrukts würdig sein soll:

„Fatigue is a suboptimal psychophysiological condition caused by exertion. The degree and dimensional character of the condition depends on the form, dynamics and context of exertion. The context of exertion is described by the value and meaning of performance to the individual; rest and sleep history; circadian effects; psychosocial factors spanning work and home life; individual traits; diet; health, fitness and other individual states; and environmental conditions. The fatigue condition results in changes in strategies or resource use such that original levels of mental processing or physical activity are maintained or reduced.“ (Phillips, 2015, S. 53).

Um weitreichendere Termini in Bezug auf Müdigkeit abzustecken und auf eine Verwendung in vorliegender Arbeit hinzuleiten, soll das Modell von May und Baldwin (2009) als Ausgangspunkt dienen (siehe Abschnitt 2.3.2). Dieses Modell beansprucht zwar primär die *Entstehung* von Müdigkeit im Fahrkontext zu erklären, arbeitet jedoch a priori mit Begrifflichkeiten, die einem logischen Verständnis des Müdigkeits-Konstrukts zuträglich sind.

2.3.2 Entstehung von Müdigkeit

In der erweiterten Fassung des von May und Baldwin (2009) vorgestellten Müdigkeitsmodells von Schmidt (2010) werden grundsätzlich drei verschiedene Ursachen unterschieden, die während der Fahrt zu einer Ermüdung führen können (vgl. Abbildung 3). Diese umfassen Überbeanspruchung, Unterbeanspruchung und schlafbezogen relevante Einflussfaktoren. Unabhängig von ihrem Ursprung resultiert jedwede Müdigkeit in einer reduzierten Leistungsfähigkeit des Menschen. Diese Annahmen sollen auch für vorliegende Arbeit Gültigkeit haben. Wichtig ist dabei zu beachten, dass sich die Termini „Über- und Unterbeanspruchung“ im Müdigkeitsmodell auf die auszuführende Aufgabe im Allgemeinen beziehen, und nicht konkret die Eigenschaften einer auszuführenden *Vigilanzaufgabe* beschreiben (wie in Abschnitt 2.2.3 verwendet). Diese Differenzierung wird im Folgenden klargestellt.

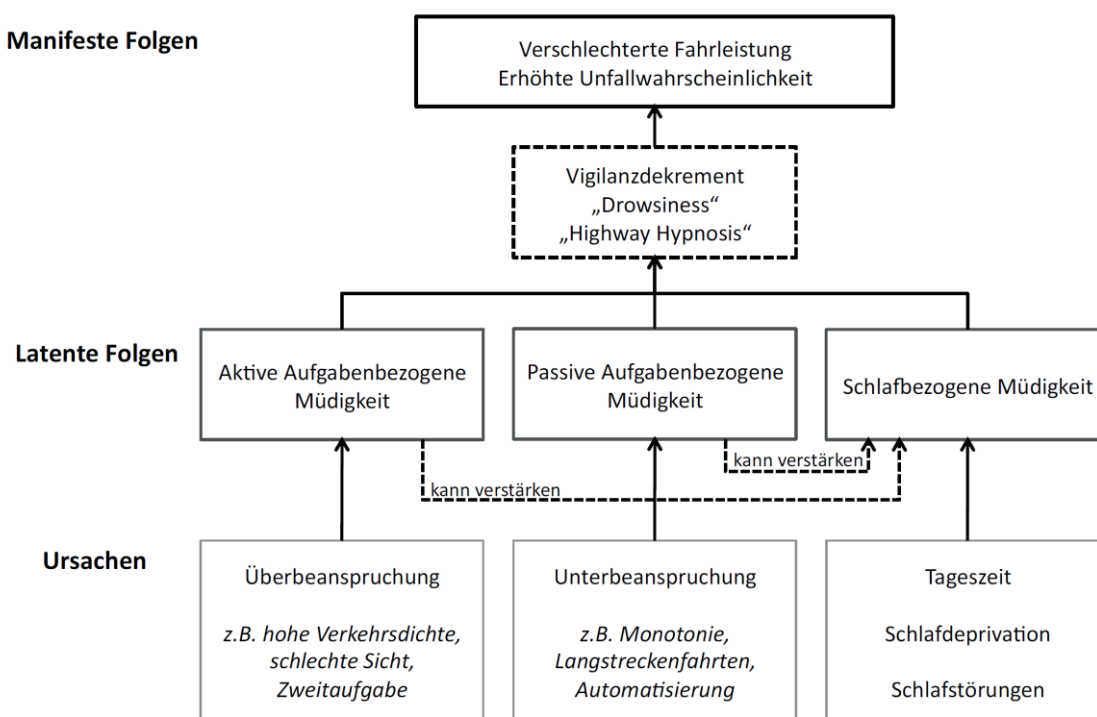


Abbildung 3: Erweitertes Müdigkeitsmodell aus Schmidt (2010), S. 8, auf Basis von May und Baldwin (2009).

Eine Überbeanspruchung kommt immer dann zum Tragen, wenn der Fahrer komplexe Tätigkeiten auszuführen hat bzw. sich in komplexen Verkehrssituationen befindet. Von dem Fahrer wird also ein hohes Maß an physischer wie psychischer Leistungsfähigkeit verlangt. Schmidt (2010) nennt hier bspw. das (manuelle) Fahren bei hoher Verkehrsdichte oder unter schlechten Sichtbedingungen (in der Nacht, bei Niederschlag oder Sichtverdeckungen). Auch unangepasste Fahrgeschwindigkeit kann die Schwierigkeit der Fahraufgabe erhöhen. Selbst ein fremdes Fahrzeug zu fahren, das sich gänzlich anders verhält als gewohnt, kann eine Überbeanspruchung bewirken. Über das prominente und viel zitierte Ressourcen-Modell von Wickens (2008) lassen sich die Arten von Überbeanspruchung generalisieren und vorhersagen, die durch zusätzliche Beschäftigung mit fahrfremden Tätigkeiten in Erscheinung treten. Je nach Modalität, die (zusätzlich) belastet wird – hier ist visuell-räumlich, auditiv, aber auch kognitiv zu nennen, verschlechtert sich die Fahrerleistung, weil weniger Ressourcen auf der eigentlichen Fahraufgabe verbleiben können. Diese Effekte wurden in der Vergangenheit breit beforscht und sind dem Themenfeld „Ablenkung beim manuellen Fahren“ zuzuordnen (Huemer & Vollrath, 2012). Für eine Übersicht kann auch Young und Regan (2007) herangezogen werden. Hargutt (2007) erklärt, wie sich aus dieser Belastung eine (Über-) Beanspruchung entwickeln kann. Demnach stellt die zu erfüllende Aufgabe bei gegebenen Umgebungsfaktoren eine spezifische Belastung für den Fahrer dar. Erst unter Berücksichtigung der spezifischen Leistungsfähigkeit des Fahrers definiert sich seine Beanspruchung. Sobald die Belastung die Fähigkeiten des Fahrers übersteigt, tritt Überbeanspruchung auf. Die Konsequenz ist, dass der Fahrer ermüdet. Diese sich dann ausbildende Müdigkeit wird nach May und Baldwin (2009) als „aktive aufgabenbezogene Müdigkeit“ bezeichnet.

Sowohl Hargutt (2007) als auch May und Baldwin (2009) beschreiben in ihren Modellen, dass neben einer Überbeanspruchung auch eine Unterbeanspruchung auftreten kann. Immer dann, wenn die Belastung die Leistungsfähigkeit des Fahrers deutlich unterschreitet, wird dies als Unterbeanspruchung definiert. Schmidt (2010) führt Beispiele solcher Fälle auf. So ist einerseits eine monotone Fahrsituation zu nennen. Hier bewirkt eine Einschränkung der auftretenden Umgebungsreize eine verminderte Belastung. Andererseits stellt eine lange Beschäftigung mit ein- und derselben Aufgabe – aufgrund ausbleibender Veränderungen – eine Quelle für geringe Belastung dar. Eine langandauernde Autofahrt fällt beispielsweise in dieses Muster. Schließlich findet sich das Beispiel „Automatisierung“ in dem von Schmidt (2010) adaptierten Modell. Wie in Abschnitt 2.1 bereits deutlich wurde, kann die Automatisierung der Fahraufgabe nicht generalisiert werden. Dennoch ist wichtig zu betonen, dass sich durch Automatisierung die Fahraufgabe fundamental ändert (Gasser & Westhoff, 2012). Vom Fahrer wird je nach Funktionsumfang (lediglich) eine Überwachungstätigkeit erwartet, oder ihm wird sogar eine vollständige

Abwendung von der Fahraufgabe zugestanden. Alle zuvor genannten Aufgaben stellen typische Vertreter sog. Vigilanzaufgaben dar, wie sie im Abschnitt 2.2 ausführliche Betrachtung erfahren. Diese verlangen in langen, monotonen Situationen eine fortwährende Aufmerksamkeitsleistung vom Fahrer. Nach May und Baldwin (2009) entsteht durch die geringe Beanspruchung eine Verminderung von Wachheit, die als „passive aufgabenbezogene Müdigkeit“ bezeichnet wird: „Closely allied with vigilance, this form of fatigue develops over a number of hours of doing what appears to be nothing at all.“ (Desmond & Hancock, 2001, S. 455). Gimeno, Cerezuela und Montañés (2006) konkretisieren mit Hilfe der Modellannahmen von Fairclough (2001), dass nicht nur die geringe Beanspruchung per se, sondern auch die kognitiven Anstrengungen im Umgang mit der Monotonie, der „mental effort“, die „passive aufgabenbezogene Müdigkeit“ hervorrufen kann, vgl. Abschnitt 2.2.3.2. Daher muss festgestellt werden, dass das Müdigkeitsmodell von May und Baldwin (2009) bzw. Schmidt (2010) für die latente Ausbildung passiver aufgabenbezogener Müdigkeit bei Nutzung einer Level 2-Automatisierung zu sparsame Annahmen trifft. So bleibt das Modell einer wirkenden Unter- und Überbeanspruchung bei der Funktionsüberwachung schuldig, wo laut Modell lediglich „Unterbeanspruchung“ bei Nutzung von Automatisierung im Allgemeinen zu passiver Ermüdung führt. Hier ist anzumerken, dass bereits Davies und Parasuraman (1982) diesen Dualismus verstanden haben. Nach ihrer Erklärung wird die Fähigkeit vigilant zu bleiben, sowohl durch physische als auch mentale Ermüdung beeinträchtigt, die bei Bearbeitung der Aufgabe hervorgerufen werden kann. Das bei längerer „time-on-task“ typische Phänomen von Vigilanzaufgaben – das Vigilanzdekrement – wird hingegen sehr wohl vom Müdigkeitsmodell (May & Baldwin, 2009; Schmidt, 2010) als manifeste Folge von Automatisierung (vermittelt über die latente Folge „passive aufgabenbezogene Müdigkeit“) erkannt. Das Auftreten passiver aufgabenbezogener Müdigkeit wurde zumindest im manuellen Fahrkontext in den letzten Jahrzehnten breit beforscht (Forsman, Vila, Short, Mott & Van Dongen, 2013; Reyner & Horne, 1998; Schmidt, 2010). Die teilweise speziell zu diesem Zweck entwickelten Messmethoden, die ebenfalls beim automatisierten Fahren ihre Anwendung finden können, werden in Abschnitt 2.3.5 vorgestellt.

Neben der zuvor thematisierten aufgabenbezogenen Müdigkeit kann sog. „schlafbezogene Müdigkeit“ in Erscheinung treten. Diese Form von Müdigkeit verläuft unabhängig von der zu leistenden Fahraufgabe: Die Ursache liegt im Schlafdrang, den jeder Mensch periodisch erlebt (Birbaumer & Schmidt, 2010). Innerhalb dieses physiologischen Prozesses kommt dem sog. Zirkadianen Rhythmus eines Menschen eine entscheidende Rolle zu (vgl. Abschnitt 2.3.3.1). Daher wird sich hier der von Dinges (1995) präferierten Bezeichnung der „sleepiness“, d.h. der „Schläfrigkeit“ als Merkmal schlafbezogener Müdigkeit angeschlossen. Ein weiteres Indiz für die Abgrenzung zwischen den ursächlichen Pfaden der aufgabenbezogenen Müdigkeit und der

schlafbezogenen Müdigkeit – Reinprecht (2013) spricht sogar von „schläfrigkeitsbedingter Müdigkeit“ – findet sich in der Kompensation dieser verschiedenen Müdigkeitstypen. Schlafbezogene Müdigkeit lässt sich ausschließlich durch Schlaf kompensieren (Hargutt, 2007). Für die Reduktion aufgabenbezogener Müdigkeit muss die Strategie bei Aufgabenbearbeitung oder die Aufgabe selbst angepasst werden. Regelmäßige Pausen helfen vor allem „aktive aufgabenbezogene Müdigkeit“ zu verringern, hingegen benötigt die Kompensation „passiver aufgabenbezogener Müdigkeit“ einen Aufgabenwechsel (Hargutt, 2007; vgl. Abschnitt 2.3.2).

Einigkeit besteht in der Forschungsgemeinschaft darin, dass „müde zu werden“, d.h. Müdigkeit zu entwickeln, bedeutet, einen Ermüdungsprozess zu durchlaufen, der Wachheit reduziert. Für das Verständnis des Konstrukts „Müdigkeit“ kann daher geschlussfolgert werden, dass die Müdigkeit eines Menschen zu einem bestimmten Zeitpunkt immer nur einen beliebig kleinen Ausschnitt des gesamten Ermüdungsprozesses darstellt. Dieser Prozess besitzt sowohl eine physiologische Komponente (schlafbezogene Ermüdung) als auch eine psychosoziale Komponente (aufgabenbezogene Ermüdung) (Aaronson et al., 1999; Brown, 1995; Stern, Boyer & Schroeder, 1994; Vogelpohl, Vollrath & Kühn, 2017). Auch Desmond und Hancock (2001) verweisen auf die multimodalen Eigenschaften von Müdigkeit, die immer dann in Erscheinung tritt, wenn diese Einflussfaktoren zusammentreffen. Konkret besitzt aufgabenbezogene Müdigkeit die Eigenschaft, schlafbezogene Müdigkeit zu verstärken (May & Baldwin, 2009). Zusätzlich zu der Ermüdung bedingenden Aufgabe – hierzu zählt also auch die Überwachung einer automatisierten Fahrfunktion gemäß SAE Level 2 – ist der Fahrer somit im Fahrzeug immer einem gewissen Maß an schlafbezogener Müdigkeit ausgesetzt. Daher werden die theoretischen Hintergründe dieses Müdigkeitstyps im nachfolgenden Abschnitt genauer betrachtet.

2.3.3 Schlafbezogene Müdigkeit

Menschen müssen aus verschiedenen Gründen schlafen, wobei zwei Theorien die breiteste Fürsprache erhalten: Nach der regenerativen Theorie nutzt der Organismus die Schlafphase, um „...das Energieniveau, das während des Wachseins abfällt, wiederherzustellen [...]“ (Pinel, Barnes & Pauli, 2019, S. 460), und um Gedächtnisinhalte zu konsolidieren (Hartmann, 1973). Xie et al. (2013) vertreten beispielsweise die Ansicht, dass die Funktion des Schlafens darin besteht, eine metabolische Homöostase im Organismus zurück zu erlangen. Sehr anschaulich vergleichen Nedergaard und Goldman (2016) den Vorgang während des Schlafens im Gehirn mit der Arbeitsweise einer Kläranlage: Das neurotoxische Peptid „Beta-Amyloid“, welches als Ablagerung im Gehirn von Alzheimerpatienten nachgewiesen wurde (Masters et al., 1985), wird während des Schlafens abgebaut. Generell nimmt die regenerative Theorie an, dass der

Zustand der Müdigkeit immer dann eintritt, wenn eine Abweichung von der Homöostase während des Wachseins besteht und folglich erst durch Schlaf und der Wiedereinstellung der Homöostase beendet ist (Pinel et al., 2019). Die adaptive Theorie des Schlafs hingegen beschäftigt sich vornehmlich mit der Frage nach Zeitpunkt und Dauer des Schlafs (Webb, 1988), statt die Funktion zu betrachten. Hiernach besitzt die „innere Uhr“ des Menschen einen etwa 24-stündigen Taktgeber, der den sogenannten Zirkadianen Rhythmus darstellt. Das impliziert, dass „...Menschen darauf programmiert sind, nachts zu schlafen, unabhängig davon, was sich tagsüber ereignet hat.“ (Pinel et al., 2019, S. 460).

2.3.3.1 Zirkadianer Rhythmus

Im Zwei-Prozess-Modell der Schlafregulation (Borbély, 1982) werden beide zuvor genannten Theorien (vgl. Abschnitt 2.3.3) eingebettet. Die homöostatische Regulation wird dabei durch den Prozess S beschrieben, der nach Borbély (1982) das optimale Verhältnis zwischen Wachheitsdauer und Schlafintensität darstellt. Sobald eine längere Wachperiode durchlebt wird, wird hierdurch typischerweise eine intensivere Schlafperiode als Ausgleich benötigt. Die Stärke von Prozess S ist demnach beschrieben als eine Funktion der vorherigen Wachheitsdauer. Der Zirkadiane Rhythmus fungiert als zweiter Prozess, Prozess C. Er wird beschrieben durch eine sinusförmige Oszillation, die nach Borbély (1982) unabhängig von den tatsächlichen Wach- oder Schlafperioden eines Menschen arbeitet. Allerdings wird von Zulley und Hajak (2005) auch auf die Abhängigkeit beider Prozesse verwiesen, wonach eine Regeneration durch Schlaf erst dann optimal ist, wenn der Zirkadiane Rhythmus dies auch vorgibt. Prozess C ist dabei endogen gesteuert, sodass nicht Umgebungsreize, sondern ein interner Zeitgebermechanismus für die Prozesstaktung verantwortlich ist. Diese Annahme wird vor allem durch die Forschungsergebnisse zu freilaufenden Zirkadianen Rhythmen untermauert. Ein freilaufender Zirkadianer Rhythmus ist nur unter Laborbedingungen, d.h. ohne die Synchronisation durch Tageslicht oder Wissen über die aktuelle Uhrzeit, zu beobachten. In diesem Zusammenhang erlangte die Versuchsreihe von Aschoff (1965) über die Forschungsgemeinschaft hinaus große Berühmtheit, in der erstmalig Probanden unabhängig von äußeren Zeitgebern mehrere Wochen in Bunker-Wohnungen lebten. Es zeigte sich, dass in den seltensten Fällen eine Taktung von exakt 24 Stunden bei einem Versuchsteilnehmer vorlag, sondern sich die Periodendauer der Circadianen Rhythmik häufig bis in den Bereich eines 25-stündigen Rhythmus verlängerte. Auch in späteren Veröffentlichungen wird eine längere Periodendauer als 24 Stunden berichtet (Czeisler et al., 1999; Wever, 2013), wobei die Periodenlänge mit der phasisch schwankenden Körperkerntemperatur einhergeht (Czeisler et al., 1999; Kerkhof & Van Dongen, 1996). Im Speziellen wird durch die

Messungen von Kerkhof und Van Dongen (1996) deutlich, dass ein Tiefpunkt der Körperkern-temperatur – und damit auch der Müdigkeit bzw. Wachheit – zwischen 03:00 Uhr und 06:00 Uhr vorherrscht; geringer fällt dieser Effekt für den Zeitbereich zwischen 14:00 Uhr und 15:00 Uhr aus. Der sog. „post-lunch dip“ ist dem Zeitbereich am frühen Nachmittag zuzuordnen. Reyner, Wells, Mortlock und Horne (2012) konnten zumindest für manuelles, monotones Fahren im Simulator mittels EEG zeigen, dass die schlafbezogene Müdigkeit signifikant mit der vorherigen Kalorienzufuhr ansteigt.

2.3.3.2 Chronotypen

Die freilaufenden Rhythmen (vgl. Abschnitt 2.3.3.1) sind zwar intraindividuell sehr konstant, weisen jedoch interindividuelle Unterschiede auf. Hierfür führt Niederl (2008) das Vorhandensein unterschiedlicher Chronotypen bei Menschen als Erklärung an. Der Chronotyp eines Menschen stellt ein zeitlich stabiles Persönlichkeitsmerkmal, einen Trait dar (Nguyen, 2018; Preckel, Lipnevich, Schneider & Roberts, 2011; Roenneberg, 2012). Gewöhnlich wird der Morgentyp (Frühtyp; auch „Lerche“ genannt) vom Abendtyp (Spättyp; auch „Eule“ genannt) unterschieden. Der Morgentyp erreicht früh sein maximales Aktivitätsniveau, wird hingegen früher müde. Dementsprechend steht der Morgentyp früh auf und geht früh schlafen – zumindest, wenn keine soziale Norm oder Arbeitszeiten (z.B. Schichtarbeit) dies beeinflussen (Wittmann, Dinich, Merrow & Roenneberg, 2006). Der größte subjektive Affekt ist nach Jankowski und Ciarkowska (2008) zwischen 09:30 Uhr und 12:00 Uhr zu beobachten. Im Gegensatz dazu benötigt der Abendtyp morgens länger, um sein maximales Aktivitätsniveau zu erreichen, wird dafür aber erst später müde, sodass er später aufsteht und später schlafen geht (externe Einflussfaktoren ausgeschlossen). Beim Spättyp berichten Jankowski und Ciarkowska (2008) hingegen, dass der größte subjektive Affekt erst ab 20:00 Uhr auftritt. Trotz großer zeitlicher Unterschiede im subjektiven Affekt, variiert die Spannweite der Bettgehzeiten zwischen den Extremtypen im Bereich von einer bis zwei Stunden (Niederl, 2008). Wittmann et al. (2006) stellten auf Basis einer großen Stichprobe ($N = 501$) dar, dass der Chronotyp in der Population normalverteilt ist, sodass die Auftretenswahrscheinlichkeit extremer Früh- oder Spättypen gering ist. Adan et al. (2012) hingegen schätzen die Wahrscheinlichkeit höher, nämlich dass ca. 20 % der Population einer der Extremtypen zugeordnet werden kann. Parallel zu dem Chronotyp haben Wittmann et al. (2006) auch den Konsum von Stimulantien in ihrer Stichprobe erhoben. Demnach zeigte sich, dass sowohl Alkohol, Tabak als auch Koffein signifikant häufiger von Spättypen als von Frühtypen konsumiert werden. Insbesondere wurden hierbei die größten Korrelationen für die jüngste Altersgruppe (14 bis 25 Jahre) ermittelt. Koffein stellt die mit Abstand am häufigsten

konsumierte psychoaktive Substanz dar (Weiß, 2007). Großer Konsens besteht in der Forschungsgemeinschaft darin, dass Koffein wachheitsfördernd und müdigkeitsunterdrückend wirkt (Smith, 2002; Zwyghuizen-Doorenbos, Roehrs, Lipschutz, Timms & Roth, 1990). Smith (2002) berichtet in seinem Review ebenfalls, dass die Performanz in Vigilanzaufgaben durch die Einnahme von Koffein gesteigert werden kann. Generell ist für die Bewertung solcher Effekte zu beachten, dass die Halbwertszeit für den Abbau im Organismus zwischen drei und sieben Stunden beträgt (Weiß, 2007).

Letztlich beschreibt Roenneberg (2012) ein Zusammenwirken verschiedener Merkmale (wie z.B. Schlafdauer- und Zeitpunkt, sowie Körperkerntemperatur), die den individuellen Chronotyp eines Menschen und damit seinen spezifischen Zirkadianen Rhythmus taktet. Für die experimentelle Anwendungsforschung im Bereich des automatisierten Fahrens definieren Körber und Bengler (2014), dass der Chronotyp des Fahrers, sowie seine vornächtliche Schlafqualität, entscheidende, schlafbezogene Faktoren darstellen, die die individuelle Müdigkeit moderieren können.

2.3.4 Müdigkeitsmodell für SAE Level 2

Im Kontext des automatisierten Fahrens wurde bis hierhin deutlich, dass das zuvor vorgestellte Müdigkeitsmodell (May & Baldwin, 2009; Schmidt, 2010) nicht mehr allen theoretischen Anforderungen an die veränderte Fahrerrolle genügt. Um den neuen Fahraufgaben in einem neuen Verkehrsszenario gerecht zu werden, wurde das Müdigkeitsmodell für die Verwendung eines Level 2 spezifiziert – bei möglichst gleichwertiger Sparsamkeit. Hierbei wurden die zuvor beschriebenen Grundsätze und Forschungsergebnisse berücksichtigt. Für das theoretische Verständnis vorliegender Arbeit werden daher die in Abbildung 4 ersichtlichen Anpassungen und Erweiterungen auf Basis von Schmidt (2010) vorgeschlagen.

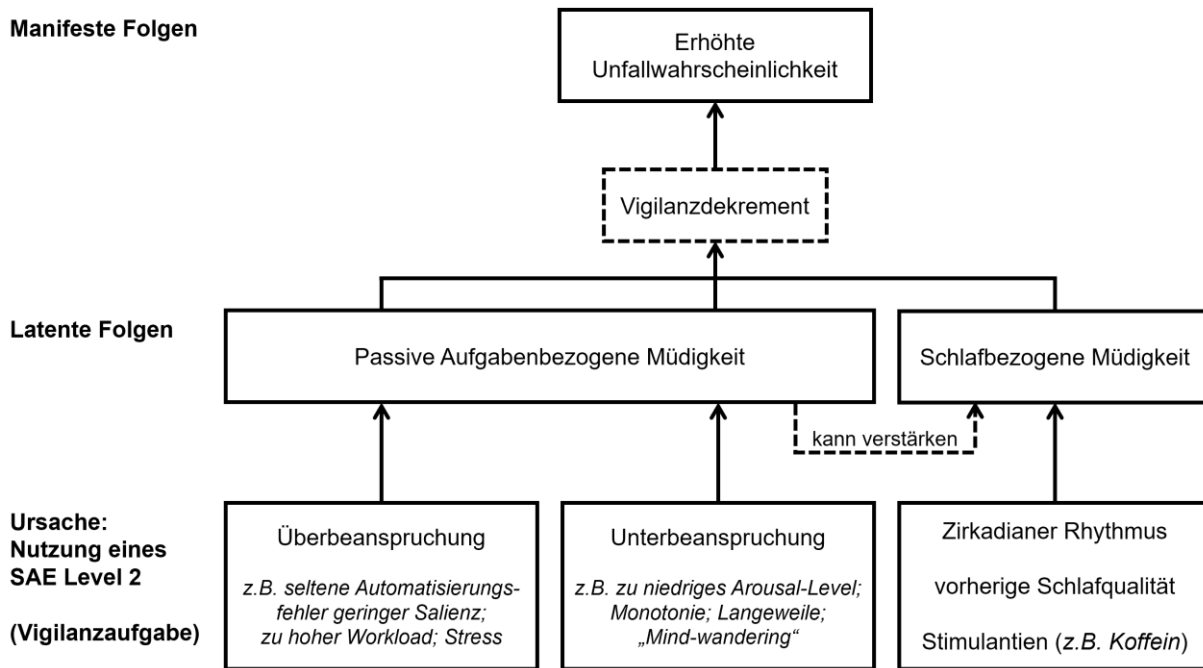


Abbildung 4: Müdigkeitsmodell für die Nutzung eines SAE Level 2, auf Basis von Schmidt (2010).

Anhand des Modells wird einerseits deutlich, dass der schlafbezogenen Müdigkeit ein nicht zu unterschätzender Beitrag für die Ausbildung des Vigilanzdekrements zukommt. Die entstehende passive aufgabenbezogene Müdigkeit macht andererseits aufgrund von parallel auftretender Über- und Unterbeanspruchung die Besonderheit bei Nutzung eines Level 2 aus. Aus diesem Grund soll sich in Abschnitt 2.4 vertieft der Auswirkung passiver Ermüdung auf den Fahrerzustand beim automatisierten Fahren gewidmet werden – jedoch mögliche schlafbezogenen Einflussfaktoren nicht außer Acht gelassen werden. Zuvor bietet sich eine Übersicht über möglichst valide Messmethoden zur Erfassung von Fahrermüdigkeit an.

2.3.5 Müdigkeitsindikatoren

Eine präzise Bestimmung von Fahrermüdigkeit wurde in den letzten Jahrzehnten durch eine hohe Rate müdigkeitsbedingter Verkehrsunfälle beim manuellen Fahren motiviert (vgl. Abschnitt 1). Um diese Unfalltypen zu vermeiden, müsste ein kritisches Müdigkeitsniveau frühzeitig während der Fahrt identifiziert und der Fahrer darauf hingewiesen werden. An dieser Stelle setzen Müdigkeitsindikatoren im Fahrkontext an. Es wird das Ziel verfolgt, den aktuellen Müdigkeitszustand des Fahrers möglichst valide und effizient abzubilden, um dann – wenn nötig – den Fahrer vor seiner nachlassenden Performanz zu warnen und eine Handlungsalternative (eine Pause) aufzuzeigen (Yan, Kuo, Lin & Liao, 2016). Dieses Wirkprinzip haben spezielle

Fahrerassistenzsysteme beim manuellen Fahren gemein, um möglichst viele müdigkeitsbedingte Verkehrsunfälle zu vermeiden (Reinprecht, 2013). Bevor ein solches Assistenzsystem Serienreife erlangen konnte, waren verschiedene, teils explorative Studien notwendig, um praktikable Müdigkeitsindikatoren untereinander vergleichend zu validieren und den bestmöglichen Indikator zu identifizieren.

Grundidee von Müdigkeitsdetektion ist es somit, das latente Konstrukt – den Fahrerzustand – mit Hilfe verschiedener (teils experimenteller) Indikatoren über manifeste Verhaltensweisen bzw. Messgrößen quantifizierbar zu machen. Einige Verfahren beanspruchen Müdigkeit indirekt zu messen, indem sie Performanz in einer monotonen Fahraufgabe oder Überwachungsaufgabe (Vigilanzaufgabe) erfassen. Daher wird im Fahrerzustandsmodell, das im Rahmen des EU-Forschungsvorhabens „Highly automated vehicles for intelligent transport“ (HAVEit; Hoeger et al., 2011) aufgestellt wurde, die Messung des Fahrerzustands grundlegend in „alertness“ und „attention“ aufgeteilt (vgl. Abbildung 5). Das Modell kommt jedoch spätestens bei Verwendung eines Level 3 an seine Grenzen, da die Erfassung der Aufmerksamkeit über Fahrerablenkung aufgrund der in Level 3 definierten Fahreraufgabe – insbesondere i.S. seiner Abwendungsbefugnis – keinen zielführenden Ansatz mehr darstellt.

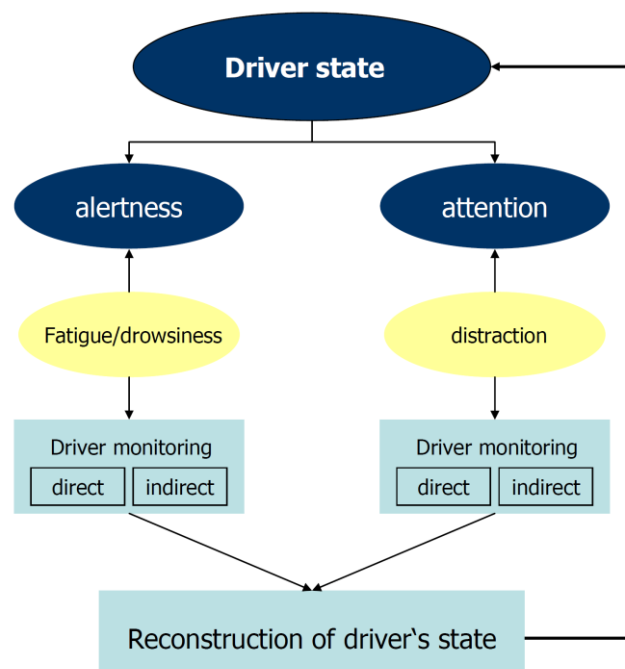


Abbildung 5: Fahrerzustandsmodell aus HAVEit (Hoeger et al., 2011, S. 105).

Eine umfangreiche Evaluation der Detektionsverfahren wurde im Auftrag der BASt von Platho, Pietrek und Kolrep (2013) durchgeführt. Dieser Bericht bietet einen Überblick über die Testgütekriterien sowie anwendungsbezogene Vor- und Nachteile der Verfahren (auch auf Basis von Expertenurteilen), die die Autoren in drei Hauptgruppen einteilen: Selbsteinschätzung, (Fahr-) Verhalten und Physiologie. Die Selbsteinschätzung des Fahrers ist ein subjektives Messverfahren, das auf die Erfassung der momentanen Müdigkeit mittels Testinstrumenten (Fragebögen) abzielt. Hingegen beanspruchen Fahrperformanzmessungen (bspw. Lenkverhalten oder Spurhaltung) sowie Verhaltensbeobachtungen (bspw. Körperbewegungen oder Lidschlüsse) aber auch physiologische Messverfahren (bspw. EEG, EOG oder EKG) das momentane Müdigkeitsniveau objektiv abzubilden. In den folgenden Abschnitten sollen die am häufigsten verwendeten Verfahren vorgestellt und hinsichtlich ihrer Validität eingeordnet werden. Hier ist vorwegzunehmen, dass gegenwärtig kein zufriedenstellender Goldstandard für die Müdigkeitserfassung im Fahrkontext existiert (Desmond & Hancock, 2001), weil Uneinigkeit darüber besteht, was als valides Außenkriterium dienen soll (Platho et al., 2013). Diese Problematik wurde bereits bei Definition des Konstrukts Müdigkeit deutlich, da unklar ist, was im besten Fall gemessen werden sollte, vgl. Abschnitt 2.3.1.

2.3.5.1 Selbsteinschätzung

Die subjektive Selbsteinschätzung der Müdigkeit zu einem beliebigen Zeitpunkt baut darauf, dass Fahrer a) in der Lage sind, durch Introspektion ihren Aktiviertheitszustand abzurufen und b) dazu fähig sind, diesen dann in quantifizierbare Aussagen zu übersetzen. Traditionelle Testinstrumente wie die „Stanford Sleepiness Scale“ (SSS; Hoddes, Zarcone, Smythe, Phillips & Dement, 1973) oder die „Karolinska Sleepiness Scale“ (KSS; Åkerstedt & Gillberg, 1990) arbeiten mit diesen Annahmen. Sie geben sieben (SSS) bzw. neun (KSS) Stufen vor, die bei der KSS von „extrem wach“ über bspw. „schläfrig, aber keine Anstrengung wach zu bleiben“ bis zu „extrem schläfrig, nicht möglich wach zu bleiben“ reichen. Obwohl Kecklund und Åkerstedt (1993) zeigten, dass die Validität der KSS – angelehnt an das objektive Außenkriterium der EEG-Alphaband-Aktivität (vgl. Abschnitt 2.3.5.4) – verhältnismäßig hoch ist, bleiben in der Forschungsgemeinschaft Zweifel, ob Menschen überhaupt zu einer objektiven Beurteilung ihres Müdigkeitsniveaus in der Lage sind. Dinges, Mallis, Maislin und Powell (1998) erklären, dass Menschen zwar generell zur Wahrnehmung von aufkommender Müdigkeit fähig sind, jedoch keine prospektiven Ableitungen über ihr Verhalten oder ihre Performanz treffen können. Schmidt et al. (2009) konnten zeigen, dass Diskrepanzen zwischen subjektiver Einschätzung der eigenen Müdigkeit und physiologischer Erfassung mittels EEG und EKG (Elektrokardiogramm) bestehen. In einer monotonen manuellen Fahrt über vier Stunden auf der Autobahn

wurden vier zeitlich äquivalente Messzeitpunkte definiert. Die mittels KSS gemessene Müdigkeit stieg monoton innerhalb der ersten drei Messzeitpunkte signifikant an, fiel jedoch bei Messzeitpunkt vier unter das Niveau von Zeitpunkt zwei ab. Hingegen wies die berechnete EEG-Alpha-Spindelrate (vgl. hierzu Abschnitt 2.3.5.4) und die abnehmende Herzrate auf einen fortwährenden, signifikanten Müdigkeitsanstieg über alle Messzeitpunkte hin. Schmidt et al. (2009) diskutieren, dass womöglich die freudige Erwartung auf das Ende der Versuchsfahrt zu dieser subjektiven Verzerrung geführt hat. Folglich können Menschen ihre Müdigkeit nicht adäquat quantifizieren. Dieses Ergebnis steht im Einklang mit dem von Di Stasi et al. (2012) gefundenen Effekt, wonach sich die mittels SSS gemessene Müdigkeit in einer simulierten manuellen Fahrt nach einer 15-minütigen Pause erholte, jedoch das objektive Außenkriterium – die Spitzengeschwindigkeit von sakkadischen Augenbewegungen – auf eine fortwährende Ermüdung hindeutete.

Darüber hinaus können verbale Müdigkeitsabfragen per se die tatsächliche (objektiv mittels EEG gemessene) Müdigkeit beeinflussen (Schmidt, Schrauf, Simon, Buchner & Kincses, 2011). Hiermit geht ein fundamentales methodisches Problem einher: Sobald Ermüdung als Kontinuum über die Zeit abgebildet werden soll, bedingen verbale Abfragen, dass kurzzeitig die objektive (mutmaßlich auch subjektive Müdigkeit) abnimmt, weil Probanden dadurch in ihrer Ermüdung „gestört“ werden. Dieser Effekt kann für etwa zwei Minuten anhalten (Schmidt et al., 2011). Auf diese Weise sind Messfehler und eine generelle Unterschätzung des tatsächlichen Müdigkeitsniveaus wahrscheinlich. Zwei weitere Einflussfaktoren – mit vordergründig untergeordneter Bedeutung – sollten für die Verfälschung subjektiver Urteile ebenfalls in Betracht gezogen werden. Einerseits ist von einer reduzierten Selbstkontrolle des Fahrers bei zunehmender Müdigkeit auszugehen, wodurch die Wahrscheinlichkeit zu lügen (paradoxerweise dann auch den eigenen Müdigkeitszustand betreffend) steigt (Mead, Baumeister, Gino, Schweitzer & Ariely, 2009). Andererseits können bei der Frage nach Müdigkeit evolutionsbiologische, negativ konnotierte Assoziationen zu Schwäche und Krankheit getriggert werden (Weinman, Petrie, Moss-Morris & Horne, 1996), die in einer leistungsorientierten Gesellschaft keine normative Konformität besitzen (Cialdini, Kallgren & Reno, 1991).

Reinprecht (2013) hat von vornherein ein fragebogenbasiertes Testinstrument entwickelt, das beansprucht, sowohl die subjektive Facette als auch die objektiven Merkmale von Fahrer-müdigkeit zu erfassen. Mit Hilfe der „Subjektiven und Objektiven Müdigkeitsskala von Braunschweig“ (SOMS; Reinprecht, 2013) ist es möglich, auf gleichem Skalenniveau die Selbstbeurteilung des Fahrers mit der Fremdbeurteilung eines sog. „Raters“ zu korrelieren. Auf diese

Weise kann Redundanzgewinn zwischen zwei Beurteilungen erzielt werden, da sich die subjektive Bewertung an einem objektiven Außenkriterium (und natürlich umgekehrt) anlehnen kann. Reinprecht (2013) berichtet einen signifikanten Zusammenhang zwischen Selbst- und Fremdreferenzierung von $r = .56$ auf Basis von $N = 1313$ Bewertungen. Das subjektive Müdigkeitsreferenzierungssystem bietet ein zweistufiges Antwortformat an: Die Müdigkeit wird mittels dreifacher Unterteilung („–“, „0“, „+“) in den Kategorien „wenig“, „mittel“, „stark“ und „sehr stark“ (Ausnahme: „wach“, keine weitere Unterteilung) erfasst. Mit Hilfe der Feingliederung der zunächst groben Antwortkategorie soll nach Reinprecht (2013) eine sehr genaue Unterscheidung innerhalb der einzelnen Müdigkeitsstadien für den Fahrer möglich sein. Das objektive Müdigkeitsreferenzierungssystem arbeitet mit nachträglicher Verhaltensbeobachtung und Kategorisierung anhand von Videodatenmaterial (siehe hierzu im Detail Abschnitt 2.3.5.3). Obwohl eine hohe Interraterreliabilität zwischen $r = .85$ und $r = .90$ von 25 Ratern bei der SOMS gegeben ist (Reinprecht, 2013), muss angemerkt werden, dass auch die objektive Fremdreferenzierung eine nach Schema und Anleitung durchgeführte (menschliche) Beurteilung darstellt. Allerdings wird hier hypothetisiert, dass die menschlichen Fehler bzw. Urteilsunterschiede gleichermaßen (vernachlässigbar) zu werten sind, wie technische Messfehler aufgrund von Messunsicherheiten oder Signalstörungen (z.B. bei EEG-Ableitung).

2.3.5.2 Fahrverhalten

Mit steigender Müdigkeit erhöht sich die Wahrscheinlichkeit von Fahrfehlern (Reyner & Horne, 1998). In einer Fahrsimulatorstudie konnten Reyner und Horne (1998) zeigen, dass „Abdriften“ aus dem eigenen Fahrstreifen unter Schlafdeprivation mit zunehmender Selbsteinschätzung der Müdigkeit (hier mittels KSS) ansteigt. Insofern ist es möglich, anhand von Lenkverhalten und Spurhaltung des Fahrers auf seinen aktuellen Müdigkeitszustand zu schließen. Damit diese Metriken den Fahrerstand valide abbilden können, wird empfohlen, spezifische Restriktionen bei der Analyse zu beachten. So ist eine hohe experimentelle Kontrolle des Verkehrsszenarios, insbesondere des Kurvenradius (bestmöglich groß zu wählen) von Vorteil (Östlund et al., 2004). In zwei breit angelegten Versuchsreihen im Fahrsimulator verglichen Forsman et al. (2013) insgesamt 87 verschiedene Metriken zur Detektion moderater Müdigkeit in 30-minütigen manuellen Fahrten bei ungefähr 90 km/h. Hierbei erwies sich die „Standard Deviation of Lane Position“ (SDLP; d.h. die Variabilität der lateralen Spurposition) als Metrik der Wahl, gefolgt von der Varianz des Lenkwinkels. Die Autoren stellen aber in Aussicht, dass speziell bei fehlenden Fahrstreifenmarkierungen oder bei schlechten Witterungsbedingungen, die Varianz im Lenkwinkel eine höhere Zuverlässigkeit besäße, als dies mittels SDLP zu realisieren ist. Interessanterweise war es Forsman et al. (2013) möglich, die zunächst auf zehn monotonen,

geraden Teilstücken erzielten Ergebnisse in einem standardisierten Testszenario mit acht kurvigen Segmenten zu replizieren. Die Länge der Geraden und Kurven betrug dabei jeweils ca. 800 m, vier der Kurven (ausschließlich Rechtskurven) waren 90°-Kurven. Hier sollte berücksichtigt werden, dass die Kurvenfahrten um kleine, zögerliche Lenkbewegungen mit Hilfe eines Hochpass-Filters bereinigt wurden. Da die Kurven verhältnismäßig eng und im Fahrsimulator durchfahren wurden, besteht Unklarheit, inwiefern sich diese Ergebnisse auf reale Bedingungen übertragen ließen.

Platho et al. (2013) erklären, dass Streuungsmaße (z.B. die Varianz) große Vorteile gegenüber generellen Abweichungen von der Fahrstreifenmitte haben, da Fahrer nicht nur aufgrund von Müdigkeit nicht in der Mitte des Fahrstreifens fahren, sondern auch Gegenverkehr hierbei eine Rolle spielen kann. Anzumerken ist darüber hinaus, dass Gewöhnung an ein häufig fremdes Versuchsfahrzeug oder bauartbedingte Verzerrungseffekte der Spurmittenzentrierung in Fahrsimulatoren (Frey, 2016) ebenfalls zu einem systematischen Versatz im Fahrstreifen führen können. Longitudinale Metriken in der Fahrzeugführung, wie z.B. die Geschwindigkeit oder Abstände zu einem vorausfahrenden Fahrzeug (Johansson et al., 2004), werden hingegen sehr selten für Müdigkeitsbewertungen herangezogen, da sie nur eine geringe Validität besitzen (Platho et al., 2013).

Das generelle Problem im automatisierten Fahrkontext besteht darin, dass der Fahrer während der automatisierten Fahrt keine Regelungstätigkeiten mittels Pedalerie oder Lenkrad mehr ausführt. Schlichtweg stehen diese Maße dann zur Abschätzung des Fahrerzustands nicht zur Verfügung. Sobald der Fahrer aber wieder manuell fährt (d.h. bspw. nach einer Übernahmeaufforderung) liegen wieder Metriken zur Beurteilung des Fahrverhaltens vor. Die im nächsten Abschnitt vorgestellte Verhaltensbeobachtung kann auch während der automatisierten Fahrt als probates Mittel zur Müdigkeitsabschätzung dienen.

2.3.5.3 Verhaltensbeobachtung

Die grundlegende Idee der Fremdbeurteilung des Fahrerverhaltens liegt darin, dass steigende Müdigkeit dem Fahrer gewissermaßen „angesehen“ werden kann. Der Lidschluss verlängert sich, die Augen können wegrollen, der Fahrer gähnt, oder er zeigt Aktivität im Sitz, um der aufkommenden Müdigkeit entgegenzuarbeiten (Rogé, Pebayle & Muzet, 2001; Vollrath, 2018; Wierwille & Ellsworth, 1994). Diese und noch weitere Verhaltensindikatoren, die sich im Schwerpunkt auf die Mimik des Fahrers konzentrieren, können genutzt werden, um die Stärke der Ermüdung zu klassifizieren, und damit zu objektivieren. Wierwille und Ellsworth (1994)

haben mit dem „Observer Rating of Drowsiness“ (ORD) erstmals eine validierte Rating-Skala (Likert-Skala) vorgestellt, die mit den zuvor genannten Charakteristika arbeitet. Die Spannbreite der kontinuierlichen, Visuellen Analogskala reicht von *nicht* müde mit dem Skalenwert (0), über *leicht* müde (25), *mittelgradig* müde (50), *sehr* müde (75) bis zu *extrem* müde (100). Eine Videoaufzeichnung vom Fahrer im Fahrzeug oder Fahrsimulator kann nach dem Versuch als Bewertungsgrundlage dienen. Die Autoren berichteten eine hohe Intraraterreliabilität von $r = .88$, sowie eine hohe Interraterreliabilität von $r = .81$ (Wierwille & Ellsworth, 1994). Kolrep, Rimini-Döring, Oehme und Jürgensohn (2005) konnten später mit einer siebenstufigen Müdigkeitsbewertungsskala nach ähnlichem Prinzip eine noch höhere Interraterreliabilität von $r = .90$ verzeichnen. Auf der Grundlage von Wierwille und Ellsworth (1994) entwickelten Wiegand, McClafferty, McDonald und Hanowski (2009) ein Trainingsprotokoll für videobasierte Müdigkeitsratings, das die jeweiligen Müdigkeitsstufen mit zu erfüllenden Verhaltenskriterien definiert, siehe Abbildung 6. Dem liegt die Idee zugrunde, die Analyse der Videoaufnahmen anhand zuvor definierter Kriterien zu präzisieren, indem die Rater die Bewertung im Vorhinein mit Hilfe von Beispielsequenzen geübt haben. So kann die Auswertungsobjektivität erhöht werden. Wiegand et al. (2009) erklären, dass die Verhaltenskriterien auf einer Zusammenstellung von acht Experteneinschätzungen beruhen.

Müdigkeitsstufe	Indikatoren
Nicht müde (0 – 12.49)	Normale Spannung des Gesichtes, schnelle Lidschlüsse, kurze schnelle Blicke, gelegentliche Bewegungen oder Gesten
Leicht müde (12.5 – 37.49)	Kein völlig waches Erscheinungsbild, leicht verlangsamte Blicke und Lidschlüsse, ausreichend aufmerksam
Mittelgradig müde (37.5 – 62.49)	Glasige Augen, verlangsamte Lidschlüsse, Verringerung der Spannung des Gesichtes, Reiben von Gesicht oder Augen, Kratzen, Gesicht verziehen, starre Blicke, Bewegungen zum Beispiel Sitzpositionswechsel
Stark müde (62.5 – 87.49)	Lidschlüsse von zwei, drei Sekunden oder länger, seitliches oder vertikales Wegrollen der Augen, keine genaue Fokussierung der Augen möglich, schielender Blick, Verringerung der Spannung des Gesichtes, Momente der Bewegungslosigkeit, vereinzelte ausgedehnte Bewegungen durch beispielsweise Lenkradkorrekturen oder das Neu Positionieren des Kopfes
Extrem müde (87.5 – 100)	Einschlafen, Lidschlüsse von vier Sekunden oder länger und ähnlich lange Momente der Bewegungslosigkeit, ausgedehnte Bewegungen bei dem Übergang in den Schlaf oder aus dem Schlaf heraus

Abbildung 6: Deutsche Übersetzung der Wiegand et al. (2009) zugrundeliegenden Verhaltenskriterien (hier Indikatoren) für die jeweilige Müdigkeitsstufe (aus Joos, 2018, S. 17).

Anders als bei Reinprecht (2013), wo sobald mindestens ein Kriterium einmal zu beobachten ist, die Müdigkeit des Fahrers auf der entsprechenden Stufe eingeordnet wird, wählten Wiegand et al. (2009) einen vergleichsweise liberalen Ansatz: Die Rater sollten die Kriterien eher als Richtlinie verstehen, da womöglich in den Videosequenzen auch andere, ihrerseits für relevant erachtete, Müdigkeitsmerkmale auftreten könnten – in dem Fall sollte eine bestmögliche Abschätzung erfolgen. Im Ansatz handelt sich bei diesen Beurteilungen um sog. „Expertenratings“ (Platho et al., 2013), die nach Ellsworth, Wreggit und Wierwille (1993; zit. nach Wierwille & Ellsworth, 1994) eine beachtenswerte Korrelation von $r = .57$ zur Amplitude im EEG-Alpha-Band aufweisen (vgl. Abschnitt 2.3.5.4).

Für die Messung bestimmter Verhaltensindikatoren, wie z.B. die möglichst exakte Bestimmung der Lidschlussdauer, muss der Rater die Mimik des Probanden sehr genau beobachten, was – so viel sei angemerkt – wiederum auch eine Vigilanzaufgabe darstellt. Dies ist aber insofern lohnend, als dass dieser Indikator das mutmaßlich valideste Verhaltensmerkmal darstellt (Schleicher, Galley, Briest & Galley, 2008). Für präzise Ausmessungen der Lidschlussdauer, aber auch für die Berücksichtigung verschiedenster Verhaltensmerkmale hat Vollrath (2018) ein hilfreiches Instrument an die Hand gegeben. Die Idee der Software „MUED 8.1 – Macht Beurteilung von Muedigkeit einfach!“ (Vollrath, 2018, 2018a) besteht darin, eine vordefinierte Videosequenz von einer Minute Länge, hinsichtlich typischer müdigkeitsbezogener Verhaltensweisen des Fahrers konzentriert zu betrachten und das Auftreten vorab definierter Merkmale in der Software zu speichern. Den Schwerpunkt dieser Verhaltensanalyse bildet die Häufigkeit und Länge von Lidschlüssen über die Videosequenz hinweg. Dabei betrachtet ein geschulter Rater die Videosequenz vorab in kompletter Länge, um einen ganzheitlichen Eindruck von der Sequenz zu erlangen. Anschließend richtet er seinen Fokus auf den Lidschluss des Probanden: Immer dann, wenn der Rater einen Lidschluss registriert, markiert er diesen durch Betätigung der Leertaste. Sind alle Lidschlüsse markiert, stellt MUED 8.1 ein Werkzeug zur Verfügung, um die Länge des jeweiligen Lidschlusses präzise (Frame-genau) auszumessen. Der Beginn eines Lidschlusses ist zu dem Zeitpunkt definiert, zu dem der Proband (vermeintlich) gerade noch etwas sehen könnte (meist ist dabei das Lid zur Hälfte geschlossen). Das Ende eines Lidschlusses wird konsistent zu dem Zeitpunkt definiert, zu dem der Proband (vermeintlich) soeben noch nichts sehen könnte (meist ist dabei das Lid wieder zur Hälfte geöffnet). Neben der Lidschlussanalyse spielt auch die Auftretenshäufigkeit von Verhaltensweisen, wie z.B. Gähnen, Kratzen, Selbstaktivierung (Änderung der Position im Sitz) oder Aufschrecken eine Rolle für die abschließende Beurteilung durch die Software: Auf Basis aller Eingaben des Raters bestimmt MUED 8.1 ein Müdigkeitsniveau, dessen Spektrum von Wachheit bis sehr

starke Müdigkeit reicht. Das Spektrum ist in insgesamt acht Kategorien untergliedert, vgl. Abbildung 7. Zusätzlich zu dem algorithmisch bestimmten Müdigkeitsniveau gibt der Rater nach genauer Sichtung der Videosequenz ein eigenes Urteil auf dieser Skala ab. Die berichteten Ergebnisse von Strelow (2018) deuten darauf hin, dass das abschließende Urteil des Raters – nachdem er alle Merkmale analysiert hat – eine validere Bestimmung der Probandenmüdigkeit gewährleisten könnte, als dies durch den Algorithmus der Software möglich ist. Grund hierfür ist, dass bestimmte Indikatoren (bspw. „Aufschrecken“ oder „Sekundenschlaf“) auch bei „sehr starker Müdigkeit“ nicht vorhanden sein müssen, weil dem Fahrer trotzdem die Augen permanent wegrollen. MUED 8.1 würde grundsätzlich von einer „starken Müdigkeit“ ausgehen, wobei „sehr starke Müdigkeit“ hier zutreffender wäre – die Autorin leitet entsprechende Anpassungen der zugrundeliegenden Müdigkeitsskala ab, um einer Unterschätzung des Müdigkeitsniveaus vorzubeugen (Strelow, 2018). Auch in der vorliegenden Arbeit wird mit MUED 8.1 gearbeitet (vgl. Abschnitt 3.1.7), allerdings wird sich auf Basis der zuvor genannten Erfahrungswerte auf die abschließende Müdigkeitsbeurteilung durch den Rater beschränkt.

Müdigkeits-Stufe	Müdigkeits-Indikatoren
0 - wach	Schnelle Lidschlüsse, unauffälliges Lidschlussverhalten, kontinuierliche Blickwechsel, Aufrechte Sitzposition, Schnelle Sakkaden, Lenkradhaltung „10 vor 2“
1 – leichte Müdigkeit (-)	Verlangsamer Lidschluss bis zu 0.5 Sekunde, müder Gesichtsausdruck, Gähnen, Reiben/Kratzen im Gesicht, Gesicht verziehen, Kopf schieflegen/abstützen
2 – leichte Müdigkeit (+)	
3 – mittlere Müdigkeit (-)	
4 – mittlere Müdigkeit (0)	Lange Lidschlüsse (ca. 0.5-1 Sekunde), Augen starren/glasiger Blick ohne Blinzeln (Dauer > 3 Sekunden), Strecken/Räkeln, Augen halb geöffnet
5 – mittlere Müdigkeit (+)	
6 – starke Müdigkeit (-)	Sehr lange Lidschlüsse (1-2 Sekunden), Wegrollen der Augen, Kopfnicken
7 – starke Müdigkeit (+)	
8 – sehr starke Müdigkeit	Lidschluss > 2 Sekunden, Sekundenschlaf, Aufschrecken

Abbildung 7: Müdigkeitsindikatoren mit dazugehöriger Müdigkeitsstufe (Skalenwert), die dem Müdigkeitsrating mittels MUED 8.1 zugrunde lagen (aus Vogelpohl et al., 2017, S. 30).²

Eine spezielle Bewertungsform der Lidschlussdauer stellt die sog. PERCLOS-Methodik dar. Mittels PERCLOS (d.h. „percent closure“) wird der prozentuale zeitliche Anteil eines um mindestens 80 % geschlossenen Augenlids an der Gesamtbeobachtungsdauer bestimmt (Barr, Howarth, Popkin & Carroll, 2005). In der Theorie beanspruchen „Eye-Tracking“-Verfahren, den

² © Gesamtverband der Deutschen Versicherungswirtschaft e.V., Unfallforschung der Versicherer, Berlin 2017, S. 30.

PERCLOS-Wert zu einem bestimmten Zeitpunkt der Fahrt automatisiert auszugeben. Dieses Prinzip wurde in den letzten Jahren vielfältig im Fahrkontext eingesetzt und die Detektionsalgorithmen wurden auch in internationalen Forschungsarbeiten stetig weiterentwickelt (Acioğlu & Erçelebi, 2016; Darshana, Fernando, Jayawardena, Wickramanayake & DeSilva, 2014; Grace et al., 1998; Yan et al., 2016). Die experimentelle Praxis sieht zuweilen so aus, dass einige Eye-Tracker Probleme mit der korrekten, individuellen Erfassung der Augen des Probanden haben. In dem Fall muss mit Messunsicherheiten in Verbindung mit fehlenden Aufzeichnungsepisoden gerechnet werden – in der Konsequenz wird die ursprünglich veranschlagte Stichprobengröße dezimiert (vgl. z.B. Feldhütter, Hecht & Bengler, 2018b). Das Tragen einer Brille (mit auftretenden Spiegelungen auf der Glasoberfläche) kann die Technik schließlich an ihre Leistungsgrenze bringen (Barr et al., 2005). Im Versuchsfahrzeug, das kongruent auch in dieser Form für die vorliegende Forschungsarbeit zur Anwendung kommen sollte (vgl. Abschnitt 3.1.4.2), wurden bereits große Schwierigkeiten bei Aufzeichnung und Auswertung des fest verbauten Eye-Tracking-Systems festgestellt und berichtet (Klamroth, Zerbe & Marx, 2019). Insbesondere muss in Realfahrtstudien mit sich verändernden Lichtverhältnissen und zusätzlichen Spiegelungen am Fahrerplatz gerechnet werden. Trotz Verwendung einer automatisierten PERCLOS-Berechnung ist daher im üblichen Fall eine aufwändige, manuelle Nachbearbeitung der aufgezeichneten Videosequenzen notwendig, um den PERCLOS-Wert möglichst exakt zu bestimmen. Sobald dieser vorliegt, besteht weitgehende Einigkeit, ein valides Maß zur Detektion momentaner Müdigkeit zu besitzen (Barr et al., 2005; McKinley et al., 2010). Platho et al. (2013) geben allerdings zu bedenken, dass lediglich hohe Müdigkeitslevel mittels PERCLOS valide bestimmt werden können, da der Operationalisierung (Augenlidschluss > 80 %) die Erfassung leichter Müdigkeitsformen verwehrt bleibt.

Trutschel, Sirois, Sommer, Golz und Edwards (2011) haben sich der relativen Validität von PERCLOS zu alternativen Müdigkeitsindikatoren, u.a. auch EEG, experimentell genähert. Hierfür haben sie in einem Übernacht-Experiment (23:30 Uhr bis 08:30 Uhr) die Müdigkeit von 16 Probanden in einem Fahrsimulator untersucht. In jeder der acht Versuchsstunden (eine Stunde Pause zwischen 03:30 Uhr und 04:30 Uhr) wiederholte sich eine 40-minütige manuelle Fahrt, gefolgt von subjektiven Abfragen und Leistungstests (wie z.B. der „Psychomotor Vigilance Task“; PVT; vgl. Abschnitt 2.2.2). Ziel war es, verschiedene Müdigkeitsindikatoren, wie PERCLOS und die Fusion aus EEG-Daten und Daten des Elektrookulogramms (EOG), hinsichtlich ihrer Validität zu vergleichen. Als Außenkriterium diente einerseits ein subjektives Maß, die KSS, andererseits ein objektiver Indikator, die „Variation of Lane Deviation“ (VLD), d.h. ein Streuungsmaß der Spurhaltefähigkeit des Fahrers. Die Autoren spezifizierten anhand dieser Außenkriterien, ob ein Proband in einer jeweiligen Versuchsstunde ermüdete oder nicht

ermüdete. Darauf aufbauend wurde getestet, welcher Indikator diese Gruppenzuordnung am besten präzisieren konnte. Die EEG/EOG-Messwerte erlaubten bei der Gruppenrückzuordnung einen mittleren Testfehler von 13 % verglichen mit der KSS, und einen mittleren Testfehler von 10 % verglichen mit der VLD. Hingegen zeigte sich, dass sich die Rückzuordnung in die Gruppen anhand der PERCLOS-Werte basierend auf der KSS auf einen Testfehler von 26 % signifikant verschlechterte, gegenüber der VLD lag diese Verschlechterung noch signifikant höher, bei 32 %. Die Autoren sprechen sich daher unter dem Titel „PERCLOS: An Alertness Measure of the Past“ (Trutschel et al., 2011, S. 172) dafür aus, von PERCLOS zukünftig Abstand zu nehmen, und stattdessen physiologischen Maßen in Kombination mit Fahrverhaltensindikatoren zur validen Müdigkeitsdetektion den Vorzug zu geben. Im nachfolgenden Abschnitt soll daher auf EEG-basierte Verfahren, im Schwerpunkt auf die Analyse von Aktivität im Alpha-Band, eingegangen werden, bei denen sich der Anwender jedoch auch besonderen Herausforderungen stellen muss.

2.3.5.4 EEG-Alpha-Aktivität und Alpha-Spindeln

Um verschiedene Ermüdungs- und Schlafstadien auf physiologischer Basis voneinander abgrenzen können, ist der Einsatz der Elektroenzephalografie (EEG) von essenzieller Bedeutung. Das von Hans Berger im Jahr 1929 erstmals während eines neurochirurgischen Eingriffs demonstrierte Verfahren erlaubte es, elektrische Potentiale am Neocortex des Gehirns zu messen (Berger, 1929; Tudor, Tudor & Tudor, 2005). Bereits Berger verwendete für die verschiedenen beobachteten Aktivitätsmuster die Begriffe „Alpha“- und „Beta“-Wellen (Tudor et al., 2005). Heutzutage können die Spannungsschwankungen nicht-invasiv mit Hilfe von Elektroden, die auf der Kopfoberfläche appliziert werden, erfasst werden. Die technische Herausforderung einer validen Ableitung besteht darin, die sehr geringen Spannungen in der Größenordnung von 100 μ V adäquat zu verstärken. Unter Laborbedingungen können auch etwaige Störeinflüsse (bspw. elektromagnetische Strahlung oder Bewegungen des Probanden) minimiert werden – unter realen (Versuchs-) Bedingungen – bspw. im Innenraum eines fahrenden Pkw – werden jedoch aktive Elektroden mit eigener elektromagnetischer Abschirmung zur Vermeidung von Artefakten benötigt.

Traditionell werden nach Dement und Kleitman (1957) vier Schlafstadien (bzw. Ermüdungsstadien) unterschieden. In Abbildung 8 (vgl. Pinel et al., 2018) sind diese schematisch auf Basis ihres EEG-Musters dargestellt. Im aufmerksamen Wachzustand (hier Stadium null) ist ein niedrig amplitudiges, aber hochfrequentes EEG mit vorwiegender „Beta-Aktivität“ (14 - 30 Hz) zu erkennen (Birbaumer & Schmidt, 2010; Carlson, 2004). Dieses Muster wird im entspannten

Wachzustand bis hin zum müden Wachzustand, also während eines Ermüdungsprozesses, durch sog. „Alpha-Aktivität“ unterbrochen. Diese „Alpha-Wellen“ besitzen eine Frequenz zwischen acht und 12 Hertz (Wickens et al., 2013). Ihre höhere Amplitude, die der des ersten Schlafstadiums ähnelt, können die Alpha-Wellen als „Vorboten“ des EEG-Musters in Schlafstadium 1 definieren. Dieses Frequenzband, das Alpha-Band, wird sich häufig zunutze gemacht, um sowohl die Aktiviertheit bei Vigilanzaufgaben (vgl. Abschnitt 2.2) als auch die Quantität der Ermüdung (d.h. den Übergang zwischen Wachheit und Schlaf) abzubilden. Eine spezielle Methode stellt dabei die Analyse sog. Alpha-Spindeln dar, welche im Folgenden sowie in Abschnitt 3.1.4.3 näher erläutert wird. Wenn schließlich Schlafstadium 1 erreicht ist, liegen höhere Anteile der Alpha-Aktivität aber auch sog. „Theta-Aktivität“ (3.5 - 7.5 Hz) vor (Carlson, 2004). Mit Übergang zu Schlafstadium 2 sinkt die Frequenz im EEG immer weiter ab – es ist „Delta-Aktivität“ (1 - 2 Hz) zu messen, die hier bereits 10 bis 50 % am Gesamtspektrum beträgt und in Schlafstadium 3 (Tiefschlaf) auf über 50 % ansteigt (Birbaumer & Schmidt, 2010), siehe Abbildung 8.

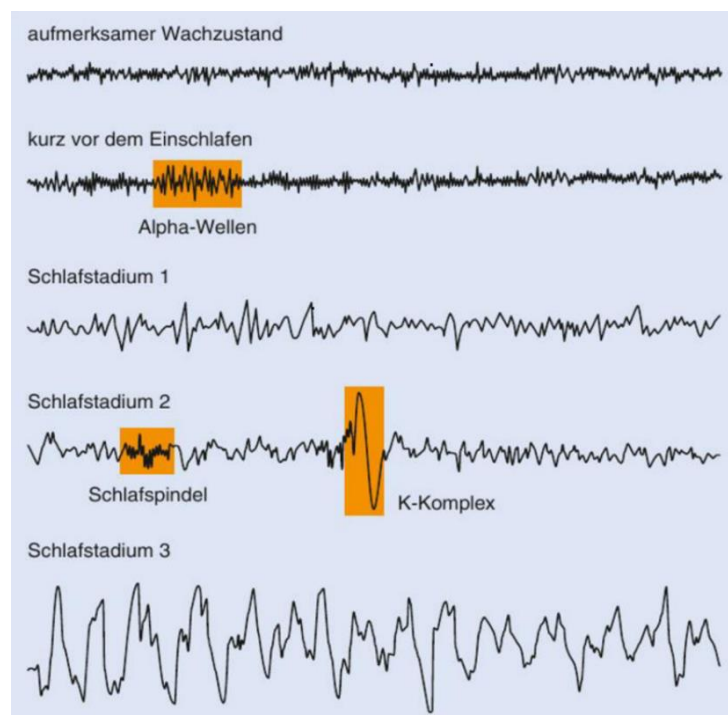


Abbildung 8: EEG-Muster abhängig vom jeweiligen Schlafstadium (aus Pinel et al., 2019, S. 457).³

Die während des Ermüdungsprozesses vorherrschende Alpha-Aktivität kann sich als objektives Maß zur Quantifizierung der Ermüdung im Fahrkontext zunutze gemacht werden (Schmidt,

³ © Pinel, John P. J.; Barnes, Steven J., Pauli, Paul; Biopsychologie, 10., aktualisierte und erweiterte Auflage, Pearson Deutschland, München 2019, S. 457.

2010). Simon et al. (2011) spezifizieren, dass vor allem in parieto-okzipitalen Regionen des Gehirns ein Anstieg von Alpha-Aktivität mit einem Anstieg von Ermüdung assoziiert ist. Diese Annahme wird durch die Beobachtungen von Van Dijk, Schoffelen, Oostenveld und Jensen (2008) gestützt, die zusätzlich Magnetoenzephalographie (MEG) zur Steigerung der räumlichen Auflösung bei der Analyse einsetzten: Probanden, die eine hohe Alpha-Aktivität in diesen Hirnregionen aufwiesen, besaßen eine signifikant niedrigere Reaktions- und Diskriminationsfähigkeit auf visuelle Reize (Van Dijk et al., 2008). Hier lässt sich ein funktionaler Zusammenhang vermuten: Der primäre visuelle Cortex, die *Area Striata*, ist in diesen okzipitalen Regionen lokalisiert – höhere visuelle Areale erstrecken sich bis in parietale Regionen (Birbaumer & Schmidt, 2010). Daher besteht die Theorie, dass eine Zunahme von Alpha-Aktivität eine aktive Inhibition in den jeweiligen Arealen zur Folge hat (Ergenoglu et al., 2004; Klimesch, Sauseng & Hanslmayr, 2007). Schmidt (2010) argumentiert daher – zumal simuliertes als auch reales (manuelles) Autofahren eine primär visuelle Aufgabe darstellt (Andersen, 2011) – dass eine objektive Analyse der Alpha-Aktivität eine zu bevorzugende Methodik für Ermüdungsmessung im Fahrkontext darstellt. Auch die Überwachung einer automatisierten Fahrt, die eine Vigilanzaufgabe darstellt (vgl. Abschnitt 2.2), ist ebenfalls als vorwiegend visuelle Tätigkeit gekennzeichnet. Torsvall und Åkerstedt (1987) berichteten daher schon früh, dass bei Zugführern, die vorwiegend visuelle Überwachungstätigkeiten ausführen müssen, eine Zunahme von Aktivität im Alpha-Band als Indikator für Schläfrigkeit bei der Arbeitsaufgabe geeignet ist. Im manuellen Fahrkontext (Autofahren) wurde sowohl in Simulatorstudien (Otmani, Pebayle, Roge & Muzet, 2005; Tietze, 2001; Xiong, Gao, Yang, Yu & Huang, 2016; Zhao, Zheng, Zhao & Liu, 2010) als auch im Realfahrzeug (Kecklund & Åkerstedt, 1993; Papadelis et al., 2007) das Alpha-Band einer Analyse unterzogen, um Fahrermüdigkeit zu detektieren.

In der in Abschnitt 2.2.3.1 vorgestellten EEG-Studie beabsichtigten Cunningham et al. (2000) das Auftreten von Tagträumen bei Vigilanzaufgaben elektrophysiologisch nachzuweisen. An dieser Stelle soll nun detaillierter auf die verwendete EEG-Methodik eingegangen werden. Der Beitrag der Studie in Bezug auf Unterbeanspruchung bei Vigilanzaufgaben wurde bereits in Abschnitt 2.2.3.1 erläutert. Innerhalb eines 30-sekündigen Fensters vor und nach einer Traumepisode wurden die EEG-Daten mit den folgenden Frequenzbandverhältnissen analysiert: a) Beta / (Alpha + Theta), b) Beta / Alpha und c) Beta / Theta. Cunningham et al. (2000) erklären, dass eine Aktiviertheitsreduktion mit einer Verringerung der Beta-Aktivität und einer Erhöhung der Alpha- sowie Theta-Aktivität einhergehen sollte (siehe hierzu auch Smit, Eling, Hopman & Coenen, 2005). Ein Anstieg des Aktiviertheitslevels hingegen sollte den umgekehrten Effekt besitzen. Entsprechend der Verhältnisse, die hierbei gebildet werden, hieße das, dass

ein hoher Quotient eine höhere Aktiviertheit, ein niedriger Quotient eine niedrigere Aktiviertheit impliziert. Daher nahmen die Autoren an, dass *vor* Ausbildung eines Tagtraums ein niedrigerer Quotient als *nach* Ausbildung eines Tagtraums zu beobachten sein müsste, da der Tagtraum aufgrund niedriger Aktiviertheit entstanden sei, jedoch das Bewusstwerden des Traums die Aktiviertheit kurzzeitig erhöht. Entsprechend der Hypothese konnte ausschließlich an parietalen Elektroden und unter Einbeziehung des Alfabandanteils (d.h. für die zuvor definierten Verhältnisse a) und b)) ein signifikant höherer Quotient nach einem Tagtraum als vor einem Tagtraum gemessen werden. Das Ergebnis verdeutlicht auch, dass Alpha-Aktivität eine valide Abbildung von Aktiviertheit (bzw. von Unterbeanspruchung) erlaubt, als dies Theta-Aktivität allein vermag. Dies steht im Einklang mit der Feststellung von Pope, Bogart und Bartolome (1995), die ebenfalls das o.g. Verhältnis a), d.h. $\text{Beta} / (\text{Alpha} + \text{Theta})$, zur Bestimmung der Aktiviertheit präferieren, mit dem Verweis, dass insbesondere eine Erhöhung der Alpha-Aktivität mit einer Verringerung der Aufmerksamkeitsleistung assoziiert ist. Bereits von Comstock, Harris und Pope (1988) wurde dies experimentell verdeutlicht, indem sie mehrere physiologische Indikatoren für Unterbeanspruchung verglichen – unter anderem auch Alpha-Aktivität.

Generell ist die Messung – und vor allem die anschließende Interpretation – der Aktivität im Alpha-Band jedoch mit großen Herausforderungen verbunden, die sich teilweise aus dem Verfahren des EEG per se ergeben. Bazanova und Vernon (2014) listen auf, dass eine alleinige Interpretation der Amplitude des Signals nicht zielführend ist, sondern probandenindividuellen Unterschieden Rechnung getragen werden sollte. Interindividuelle Varianz ergibt sich hiernach bereits aus der Impedanz bei Signalableitung: Sowohl Eigenschaften der Kopfhaut als auch die Stärke des Schädelknochens bedingen die Objektivität der Messung (Wen & Li, 2006). Unmittelbaren Einfluss auf die Alpha-Aktivität kann sich auch aus dem cerebralen Blutfluss ergeben. Eine Minderung sauerstoffreichen Blutes in okzipitalen, temporalen und frontalen Regionen resultiert in einer Erhöhung der Alpha-Aktivität (Goldman, Stern, Engel & Cohen, 2002). Es besteht hierbei die Vermutung, dass Alpha-Aktivität das (globale) Resultat kortikaler Inhibition ist, die vom Thalamus gesteuert wird (Bazanova & Vernon, 2014). Diese Folgerung deckt sich mit den Befunden von Goldman et al. (2002), die bei Zunahme von Alpha-Aktivität eine abgeschwächte Signalgebung im Thalamus feststellten. Ferner bestehen Indizien, dass ebenfalls die Ausschüttung von stressvermittelnden Hormonen (wie z.B. Cortisol oder Adrenocorticotropin (ACTH)) die Alpha-Aktivität beeinflusst, wonach bei hohem und niedrigem ACTH-Level die Alpha-Aktivität gering ist, jedoch bei moderater Konzentration das höchste Niveau erreicht (Sannita et al., 1999).

Ein spezielles Problem ergibt sich bei Menschen, deren EEG-Frequenzspektrum nicht in den Alpha-Bereich vordringt, weil ihr Spannungsniveau generell zu niedrig ist – hier ist die Rede von „low voltage EEG“ (Anokhin et al., 1992). Manchmal ist bei diesen Personen zumindest diffuse oder unterbrochene Alpha-Aktivität nachweisbar (Bazanov & Vernon, 2014). Nach Bodrov, Malkin, Pokrovskii und Shpachenko (1984; zit. nach Bazanov & Vernon, 2014) betrifft dies zwischen drei und 13 % der Bevölkerung. Betrachtet man die Erklärungsansätze, so kommen nach Bazanov und Vernon (2014) verschiedene Gründe, wie z.B. Traumata, Alkoholismus, oder ein genetischer Einfluss in Frage (siehe auch Steinlein, Anokhin, Yping, Schalt & Vogel, 1992). Insgesamt ist die Befundlage jedoch unzureichend. Fest steht jedoch, dass bei Menschen mit „low voltage EEG“ eine Müdigkeitsmessung auf Basis von Alpha-Aktivität sehr uneindeutig bis hin zu unmöglich ist. Hier kann nicht mit Sicherheit geklärt werden, ob diese Personen eine geringe Müdigkeit besitzen, oder die Müdigkeit lediglich nicht messbar ist. Für weitere interindividuelle Unterschiede und spezifische Herausforderungen bei Analyse von EEG-Alpha-Aktivität sei an dieser Stelle auf die ausführliche Darstellung von Bazanov und Vernon (2014) verwiesen.

Simon et al. (2011) haben ihre Arbeit auf dem zuvor beschriebenen Grundprinzip aufgebaut, wonach höhere Alpha-Aktivität mit höherer Müdigkeit assoziiert ist, mit dem Ziel, die Messung von Alpha-Aktivität als Ermüdungskorrelat robuster und damit valider zu gestalten. Die Autoren heben die hohe Objektivität eines EEG-basierten Verfahrens hervor, weisen allerdings auf das Problem hin, dass vor allem im Realfahrzeug erhebliche Störeinflüsse vorhanden sind, die Artefakte im EEG bedingen. Zudem ist die anschließende Auswertung und Interpretation äußerst zeitintensiv. Daher bestand die Anforderung, eine Müdigkeitsmessung mittels EEG zu entwickeln, die auch während der Fahrt im realen Straßenverkehr valide und quasi in Echtzeit automatisiert ausgewertete Müdigkeitswerte liefert. Die hierfür von Simon et al. (2011) erstmals vorgestellte und angewandte Methode der „Alpha-Spindel“-Analyse verzichtet auf die herkömmliche Betrachtung des gesamten Alpha-Bands und bedient sich stattdessen sog. Anhäufungen („Bursts“) von Aktivität im analogen, bzw. leicht erweiterten Frequenzspektrum zwischen sieben und 13 Hertz (Schmidt, 2010). Diese Spindeln stellen spektrale Mikrostrukturen dar, die sich durch eine schmale und stabile Frequenzspitze im Alpha-Band mit einer sich daran anschließenden niedrig-frequenten Hüllkurve beschreiben lassen (vgl. Abbildung 9; Simon et al., 2011). Die Detektion verläuft elektrodenspezifisch: Jede ein-sekündliche Sequenz mit Überlappung von 750 ms wird auf ein Amplitudenmaximum im Alpha-Band hin untersucht. Entsprechend der Annahmen der Autoren können Alpha-Spindeln dann über ein spezifisches Signal-Rausch-Verhältnis aufgespürt werden, wobei vorab eine Zerlegung der Frequenzanteile mittels Fourier-Transformation erforderlich ist (für weitere Berechnungsdetails

siehe Simon et al., 2011). Die Ausbreitungsdauer kann insgesamt mehrere Sekunden betragen. Eine Alpha-Spindel wird folglich als ein diskretes Ereignis im Alpha-Band verstanden, das durch seine Dauer, Amplitude und Frequenzspitze charakterisiert ist. Die Auftretenshäufigkeit dieses Ereignisses in gegebener Zeit wird als „Alpha-Spindelrate“ bezeichnet (Simon et al., 2011).

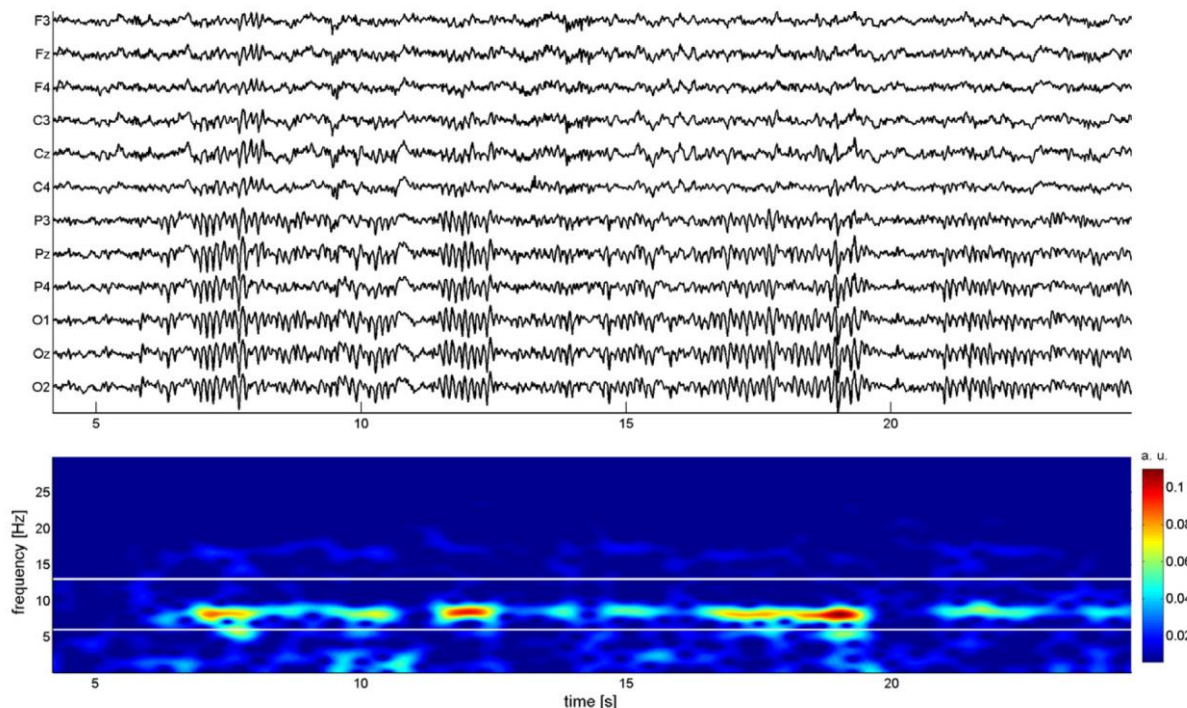


Abbildung 9: Beispielhafte Alpha-Spindel Aktivität über einen Zeitbereich von 25 s. Oben: Elektrodenpositionen (Kanäle), an denen die Ableitung erfolgte. Deutlich wird ein ausgeprägter Spindel-Charakter an parieto-okzipitalen Elektroden. Unten: Summierte Spindeldichte im Frequenzbereich zwischen 7 und 13 Hz (begrenzt durch weiße Linien) (aus Simon et al., 2011, S. 1170).⁴

Nach Schmidt (2010) besitzt die EEG-Alpha-Spindel-Methode den Anspruch, jedwede Form von Müdigkeit, unabhängig von ihrem Entstehungsmechanismus (vgl. hierzu May & Baldwin, 2009; Abschnitt 2.3.2) zu detektieren. Schmidt (2010) argumentiert, dass die Messung von Müdigkeit per se beim vorgestellten Verfahren das Ziel ist, da alle latenten Müdigkeitstypen mit einer erhöhten Unfallwahrscheinlichkeit einhergehen können. Grundsätzliche Kritik an EEG-basierten Verfahren, die auch für die Bestimmung der Alpha-Spindelrate gilt, kann dahingehend geäußert werden, dass trotz EEG-typischer hoher Abtastrate, durch Summation und

⁴ Abgedruckt aus „EEG Alpha Spindle Measures as Indicators of Driver Fatigue under Real Traffic Conditions“ von Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., ... & Schrauf, M., erschienen in „Clinical Neurophysiology, 122(6), 1168-1178“, Copyright 2011, mit Erlaubnis von Elsevier.

Mittelung der Spannungsschwankungen nur eine geringe zeitliche Auflösung für die Müdigkeitsdetektion erhalten bleibt (Platho et al., 2013). Auch in der vorliegenden Arbeit wurde die primäre (objektive) Detektion von Müdigkeit mittels EEG-Alpha-Spindelrate durchgeführt (vgl. für weitere Details Abschnitt 3.1.4.3).

2.3.6 Zusammenfassung Müdigkeit

Eine Definition von Müdigkeit gestaltet sich schwierig, da unklar ist, welche Merkmale entscheidend für eine Quantifizierbarkeit sind. Häufig werden abhängig von der Forschungsfrage angemessene Umschreibungen erzeugt, die hilfsweise mit entsprechenden Messmethoden erfasst werden können. Grundsätzlich sollte zwischen schlafbezogener und aufgabenbezogener Müdigkeit differenziert werden. Schlafbezogene Müdigkeit umfasst das physiologische, periodische Bedürfnis nach Schlaf, das durch den zirkadianen Rhythmus und den individuellen Chronotyp eines Menschen charakterisiert ist. Aufgabebezogene Müdigkeit stellt die psychosoziale Komponente des Konstrukts dar, die neben der aktiven Form (Überbeanspruchung durch hohen Workload) auch passiv (Unterbeanspruchung durch die entsprechende Aufgabe) in Erscheinung treten kann. Letztere Ausprägung ist von hoher Relevanz bei der Nutzung einer automatisierten Fahrfunktion. Obwohl der Fahrer während Level 2 zu einer permanenten Überwachungsleistung fähig sein muss oder sich während Level 3 gemäß §1b StVG „wahrnehmungsbereit“ halten soll, ist modelltheoretisch davon auszugehen, dass insbesondere in Level 2 durch die Bearbeitung einer typischen Vigilanzaufgabe passive aufgabenbezogene Müdigkeit hervorgerufen wird. Die Folge ist ein Vigilanzdekrement, was wiederum eine erhöhte Unfallwahrscheinlichkeit nach sich zieht. Aufgabenbezogene Müdigkeit besitzt zudem das Potential, schlafbezogene Müdigkeit zu verstärken.

Bereits beim manuellen Fahren ist davon auszugehen, dass womöglich 15 bis 20 % aller Verkehrsunfälle auf „Übermüdung“ zurückzuführen sind (unter Berücksichtigung der Dunkelziffer). Daher wurden verschiedene Müdigkeitsindikatoren entwickelt, um das Ermüdungsverhalten des Fahrers während der Fahrt quantifizierbar zu machen. Hierzu zählen die subjektive Selbsteinschätzung (geringste Validität), die Beurteilung des Fahrverhaltens, die (videobasierte) Müdigkeitsbeobachtung (insbesondere über Lidschlussanalyse) sowie spezielle physiologische Maße, wie EKG oder EEG. Für EEG-basierte Müdigkeitsdetektion muss zwischen der Analyse von Frequenzbandverhältnissen und kanalspezifischer Analyse diskreter Ereignisse im Alphaband (Alpha-Spindeln) unterschieden werden. Nicht alle Müdigkeitsindikatoren, die sich zuletzt in der Forschung zum manuellen Fahren bewährt haben, sind auch eins zu eins auf den automatisierten Kontext übertragbar.

Im folgenden Abschnitt soll zunächst die Erforschung des Fahrerzustands im Bereich der Mensch-Maschine-Interaktion beim automatisierten Fahren eingeordnet werden. Anschließend wird der Fokus entsprechend des in Abschnitt 2.3.4 angepassten Müdigkeitsmodells verstärkt auf anwendungsorientierte Befunde und Konzepte für Level 2 in Abgrenzung zu Level 3 gelegt.

2.4 Fahrerzustand beim automatisierten Fahren

Im Vorfeld der Literaturanalyse zum Fahrerzustand beim automatisierten Fahren empfiehlt sich ein Überblick über die gegenwärtige Publikationsdichte. Dabei soll die Nennung verschiedener Begriffe in dem Themenfeld differenziert betrachtet werden. Die Auflistung auf Basis einer Abfrage der Datenbank „PsycINFO“ der American Psychological Association (APA) in Tabelle 2 lässt darauf schließen, dass neben dem Terminus „automated driving“ – mit der mit Abstand höchsten Trefferrate in allen Feldern, Abstract und Schlüsselwörtern – auch „automated vehicle“, „vehicle automation“ und „autonomous driving“ eine wichtige Rolle spielen.

Tabelle 2: PsycINFO (APA) Abfrage vom 25.06.2020: Anzahl von Treffern für Begriffe im Themenfeld des automatisierten Fahrens.

	alle Felder	Abstract	Schlüsselwörter
„automated driving“	409	170	108
„automated vehicle“	223	58	14
„vehicle automation“	279	63	37
„autonomous driving“	239	44	27

Dem Publikationsausmaß soll sich im Folgenden am Beispiel von „automated driving“ weiter genähert werden. Für einen theoretischen Hintergrund dieser Arbeit ist vor allem experimentalspsychologische Forschung mit dem Schwerpunkt Mensch-Maschine-Interaktion von Relevanz – konkret den Fahrerzustand bei Verwendung einer automatisierten Fahrfunktion betreffend. Tabelle 3 gibt daher einen Überblick über die kombinierte Nennung von „automated driving“ mit zuvor thematisierten fahrerzustandsrelevanten Begriffen (vgl. Abschnitt 2.2 und 2.3).

Tabelle 3: PsycINFO (APA) Abfrage vom 25.06.2020: Anzahl von Treffern für Begriffe im Themenfeld des Fahrerzustands in Kombination mit „automated driving“.

	alle Felder	Abstract	Schlüsselwörter
„state“	183	14	2
„vigilance“	80	6	1
„fatigue“	103	9	3
„drowsiness“	42	4	2
„sleepiness“	36	3	0

Der globale Begriff des Zustands („state“) erzielt hier erwartungsgemäß die meisten Treffer. Auffallend ist jedoch, dass Fahrerzustand, Vigilanz und Müdigkeit häufig nicht primärer Forschungsgegenstand sind, da diese Begriffe nur selten als Schlüsselwort gewählt wurden. Ferner spiegelt sich die zuvor monierte uneinheitliche Benennung innerhalb des Müdigkeitskonstrukts wider, wobei „fatigue“ (analog zu vorliegender Arbeit) noch am häufigsten Verwendung findet. Betrachtet man darüber hinaus den zeitlichen Verlauf von Veröffentlichungen mit Nennungen von „automated driving“ in Kombination mit mindestens einem der in Tabelle 3 gelisteten fahrerzustandsbezogenen Begriffen aus allen Feldern (vgl. Abbildung 10), wird deutlich, dass der Forschungsbereich erst in den letzten Jahren sprunghaft, annähernd exponentiell gewachsen ist.

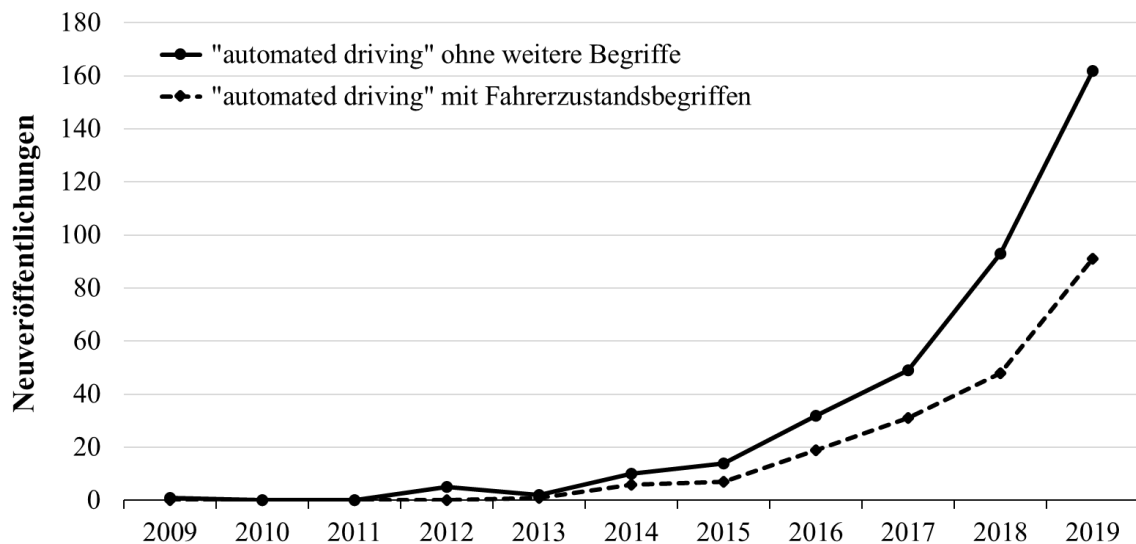


Abbildung 10: PsycINFO (APA) Abfrage vom 25.06.2020: Anzahl von Neuveröffentlichungen mit den Begriffen „automated driving“ sowie „automated driving“ in Kombination mit mindestens einem der Begriffe „state“, „vigilance“, „fatigue“, „drowsiness“ oder „sleepiness“ in allen Feldern für die Erscheinungsjahre 2009 bis 2019.

Die größte Herausforderung bei der Literaturanalyse stellt sicherlich die uneinheitliche Benennung der Ausgestaltung und des Leistungsumfangs der in Experimenten verwendeten automatisierten Fahrfunktion dar. Obwohl frühzeitig ein Level 2 oder Level 3 als solches definiert wurde (Gasser et al., 2012), orientieren sich viele Publikationen nicht streng an den festgelegten Kriterien. Die Tatsache, dass Längs- und Querführung automatisiert wurden, der Fahrer dies aber permanent überwachen musste, heißt beispielsweise nicht zwangsläufig, dass eine „voll-automatisierte“ – oder gar „autonome“ – Fahrfunktion genutzt wurde. Des Weiteren bedeutet mit einem Level 3 zu fahren nicht, dem Fahrer eine großzügige Vorlaufzeit zur Übernahme der manuellen Fahrzeugsteuerung zu gewähren, ihm aber gleichzeitig aufzuerlegen, die Arbeitsweise der Funktion permanent überwachen zu müssen. Neubauer, Matthews und Saxby (2014) bezeichnen z.B. das Fahren mit Tempomat (richtigerweise SAE Level 1) fälschlicherweise als „Partial automation“ – mit der Begründung, dass nur die Längs-, nicht aber die Querführung automatisiert werden.

In einem Großteil von Studien wird nicht das Automatisierungslevel, sondern der (vermeintlich richtige) dazu gehörige Terminus gemäß SAE-Standard J3016 (SAE, 2018) verwendet, d.h. „Partial Automation“ (Teilautomatisierung) für Level 2, „Conditional Automation“ (Bedingte Automatisierung) für Level 3. Nachsicht sollte man mit Autoren haben, die eine Level 3-Funktion als „High Automation“ (Hochautomatisierung) bezeichnen, obwohl dies ein Level 4 beschreiben würde. Tatsächlich wurde diese Benennung von Gasser et al. (2012) ursprünglich

verwendet, jedoch von der SAE im weiteren Verlauf in „Conditional Automation“ abgeändert. Aus diesem Grund wurde auch das Verbundprojekt „Kooperatives, hochautomatisiertes Fahren“ (Ko-HAF; Hohm et al., 2018), das auf Level 3 abzielte und in dessen Rahmen vorliegende Arbeit entstand, seinerzeit noch so benannt. Statt lediglich das Automatisierungslevel oder den zugehörigen Terminus für eine Einordnung der berichteten Ergebnisse im Zuge der Literaturrecherche zugrunde zu legen, ist es umso wichtiger, sowohl den experimentellen Aufbau (i.S. der Funktionsauslegung) als auch die Probandeninstruktion (i.S. der Funktionsbeschreibung) präzise zu hinterfragen.

2.4.1 Überblick über Themenfelder und Forschungsbedarf

Im Rahmen der „Arbeitsgruppe Forschung“ des vom BMVI initiierten „Runden Tisches Automatisiertes Fahren“ (RTAF) wurde umfangreicher Forschungsbedarf in Bezug auf die Mensch-Maschine-Interaktion beim automatisierten Fahren identifiziert (Gasser et al., 2015). Selbst wenn der Abschlussbericht dieser Arbeitsgruppe eingesteht, lediglich die Anfang des Jahres 2015 absehbaren Fragestellungen aufzuwerfen, so besticht er dennoch durch fortwährende Aktualität: „Der Mensch wird auch im Fall hoher Automatisierung eine wichtige Rolle im Fahrzeug spielen, um an den Systemgrenzen oder auch bei Systemfehlern die Situation zu stabilisieren.“ (Gasser et al., 2015, S. 6). Damit dies für den Menschen auch leistbar ist, setzen die Autoren die Erfassung des Fahrerzustands und der Fahrer Verfügbarkeit auf Platz eins der vorgeschlagenen (Forschungs-)Reihenfolge. Erst wenn umfassend geklärt ist, wo die Leistungsgrenze des Menschen bei Nutzung der jeweiligen Fahrfunktion liegt, kann hierauf mit ergonomischen Maßnahmen reagiert werden. Besonders interessiert dabei die sog. „Übernahmebereitschaft“ des Fahrers. Hinter dieser Forschungsfrage steht die Problematik, dass die automatisierte Fahrperiode nach einer bestimmten Zeitdauer endet und sich (ohne dazwischentretendes Anhalten) manuelles Fahren anschließt. Es geht also um die Gewährleistung eines adäquaten Fahrerzustands, um eine Übernahme zum manuellen Fahren (verkehrs-)sicher vollziehen zu können. Nur selten wird in Forschungsarbeiten die Gelegenheit ausgelassen, darauf hinzuweisen, dass automatisierte Prozesse – wie auch Fahrzeugautomatisierung – hinsichtlich der benötigten Vigilanz bzw. dem auftretenden Vigilanzdekrement kritisch zu beurteilen sind (Warm et al., 2018). Molloy und Parasuraman (1996) mahnen, dass Vigilanzprobleme eine ernstzunehmende Unfallursache darstellen. Warm et al. (2018) fordern sogar den Schritt in höhere Automatisierungslevel: Um den Menschen als Fehlerquelle bei Überwachungsaufgaben auszuschließen, hilft es die Überwachungsnotwendigkeit aufzuweichen oder gänzlich obsolet werden zu lassen.

Für Level 2 werfen Gasser et al. (2015) daher die Frage auf, ob der Fahrer überhaupt im Stande ist, eine geforderte langandauernde, kontinuierliche Überwachungsaufgabe auszuführen. Aber auch in Level 3 geht es darum, Übernahmebereitschaft zu gewährleisten, indem ein adäquater Fahrerzustand vor der (endgültigen) Übernahme sichergestellt wird. Einflüsse können sich nach Gasser et al. (2015) beispielsweise durch die Bearbeitung fahrfremder Tätigkeiten (FFT) ergeben. So stellen die Autoren die Verheißung als Frage in den Raum, ob die Beschäftigung mit bestimmten FFT einem angemessenen Fahrerzustand zuträglich sei, um den Fahrer auf einem höheren Aktivierungsniveau zu halten.

Eine notwendige Bedingung für die fahrerseitige Verwendung von FFT stellt Vertrauen in die Level 3-Funktion dar (Miglani, Diels & Terken, 2016). Hier besteht die Frage, welche Funktionscharakteristika – z.B. die Funktionszuverlässigkeit oder die Erfahrung mit der Arbeitsweise der Funktion – entscheidend für ein hinlängliches Vertrauen sind (Hergeth, Lorenz, Vilimek & Krems, 2016; Körber, Baseler & Bengler, 2018; Metz, Wörle, Hanig, Schmitt & Lutz, 2020). Ferner ist dann von Interesse, welchen fahrfremden Tätigkeiten bevorzugt von Fahrern nachgegangen werden möchte. In Hinblick auf eine hohe externe Validität im Experiment wäre es nicht zielführend, lediglich auf standardisierte (Neben-)Aufgaben aus der Ablenkungsforschung zurückzugreifen, da womöglich der motivierende und damit aktivierende Aspekt der Tätigkeit vernachlässigt wird. Dies kann insbesondere bei Erforschung des Fahrerzustands einen wichtigen Einflussfaktor darstellen. Erste Untersuchungen gehen der Evaluation von präferierten FFT bei Level 3 bereits gezielt nach (Wörle & Metz, 2020).

Mit höherem Automatisierungsgrad und durch die Bearbeitung von FFT steigt jedoch die Wahrscheinlichkeit, dass der Fahrer Kinetose (im Verkehrskontext meist als „Motion Sickness“ bezeichnet) entwickelt (Sivak & Schoettle, 2015). Eine Übersicht verschiedener Theorien, warum Kinetose entsteht, findet sich in Frey (2016) am Beispiel von Fahrsimulation. Ein grundlegendes Problem ab Level 3 besteht darin, dass der Fahrer zeitweise eine Passagier-Rolle einnimmt, und kein visuell-vestibulärer Abgleich durch Beobachtung der Fahrt mehr stattfindet (siehe z.B. Wada, 2016). Besonderes Forschungsinteresse wird diesem Bereich gewidmet, weil deutliche Hinweise auf eine psychosomatische Verknüpfung zwischen der Ausbildung von Kinetose und der Ausbildung von Müdigkeit bestehen (Berry, 2017; Gianaros, Muth, Mordkoff, Levine & Stern, 2001; Katz, 2011). Hierdurch wird unweigerlich der Fahrerzustand und damit die Übernahmebereitschaft beeinflusst.

Um eine adäquate Übernahmebereitschaft zu gewährleisten muss ebenfalls das Anzeigekonzept der Mensch-Maschine-Schnittstelle mit großer Sorgfalt gestaltet werden. Andernfalls besteht

Gefahr, sog. „Mode Confusion“ hervorzurufen: Der Fahrer muss eindeutig verstehen, in welchem Level er sich befindet. Wenn die Automatisierung deaktiviert ist, darf der Fahrer nicht davon ausgehen, noch mit aktivierter Automatisierung zu fahren. Wie ist das zu gewährleisten? Feldhütter, Härtwig, Kurpiers, Hernandez und Bengler (2018a) versuchten beispielsweise anhand visueller Aufmerksamkeitsbestimmung (mittels Eye-Tracking) Degradation von sog. „Mode Awareness“ bei Wechsel von Level 2 über Level 3 und zurück zu Level 2 sichtbar zu machen. Es zeigte sich, dass die Probanden eine signifikant niedrigere visuelle Aufmerksamkeitsleistung für die Überwachung in der zweiten Level 2-Phase verglichen mit der ersten aufbrachten. In einem anschließenden halbstrukturierten Interview stellte sich u.a. heraus, dass sich die Versuchsteilnehmer im Unklaren über die erlaubte Verwendung von fahrfremden Tätigkeiten während Level 2 waren. Statt Einbußen im „Mode Awareness“ war entsprechend der Autoren ein blindes Vertrauen in die sehr zuverlässige Level 2-Funktion für die Aufmerksamkeitsreduktion verantwortlich (Feldhütter et al., 2018a).

Insofern wird gegenwärtig an verschiedenen Gestaltungsvarianten für Übernahmeaufforderungen geforscht, um nicht nur „Mode Confusion“ zu verhindern, sondern auch ein hinreichendes Situationsbewusstsein für die anschließende manuelle Fahrzeugsteuerung sicher zu stellen (siehe z.B. De Winter, Happee, Martens & Stanton, 2014; Telpaz, Rhindress, Zelman & Tsimhoni, 2015; Walch et al., 2017; Zhang, De Winter, Varotto, Happee & Martens, 2019). Anhand der prominenten Situationsbewusstseinstheorie von Endsley (1995) wird deutlich, dass der dreistufige Prozess zum Erlangen von Situationsbewusstsein (Wahrnehmung einzelner Elemente der Situation, Verständnis der aktuellen Situation und Projektion auf den zukünftigen Status) sowohl von aufgabenspezifischen Faktoren als auch von (Fahrer-)individuellen Einflussgrößen moderiert wird. Konkret erkennt Endsley (1995), dass der durch die Übernahme der manuellen Fahrzeugsteuerung hervorgerufene Stress und der vom Fahrer benötigte Workload wichtige Einflussgrößen darstellen. Die in der Theorie zu berücksichtigenden fahrerindividuellen Mechanismen zur Informationsverarbeitung lassen sich gut mit dem Konstrukt des Fahrerzustands vereinbaren – hier besteht weiterer Forschungsbedarf, inwiefern die durch die automatisierte Fahrt hervorgerufenen müdigkeitsbedingten Leistungseinbußen die manuelle Steuerungsübernahme sowie die anschließende manuelle Fahrt beeinflussen können. Aus den in Abschnitt 2.4.5 vorgestellten Untersuchungen lassen sich bereits erste Erkenntnisse ableiten.

Insgesamt sind zur Beantwortung dieser vielschichtigen Fragen valide Forschungsmethoden vonnöten. Konkret besteht die Herausforderung, ein Maß zu finden, welches geeignet ist, den Zustand des Fahrers im automatisierten Fahrzeug zuverlässig zu detektieren. Ist dies gefunden,

kann auch die Frage nach der Unterscheidung verschiedener Fahrerzustände, die objektive Beurteilung verschiedener Übernahmezeiten oder die Prädiktion des fahrerseitigen Situationsbewusstseins angegangen werden (Gasser et al., 2015). Schließlich besteht die ernstzunehmende Problematik, wie in den Abschnitten 2.2 und 2.3 deutlich wurde, dass der Fahrer bei Überwachungstätigkeiten ermüdet. Die Falsifikation der Übertragbarkeit dieser grundlagenpsychologischen Effekte auf das automatisierte Fahren wird gegenwärtig von der Forschungsgemeinschaft ins Visier genommen. Insbesondere wird die Relevanz der Erforschung des Fahrerzustands bei Betrachtung der nachfolgend erläuterten Unfälle deutlich, über die auch in den Medien berichtet wurde.

2.4.2 Unfälle während SAE Level 2

In den USA sind in den letzten Jahren insbesondere drei tödliche Verkehrsunfälle während der Nutzung einer Level 2-Funktion bekannt geworden:

Am 07. Mai 2016 kollidierte ein Tesla Model S auf einer unbeschränkten Kreuzung eines Highways in Williston (Florida) mit einer Sattelzugmaschine. Das „Automatic Emergency Braking“ (AEB) warnte den Fahrer nicht und leitete auch keine Bremsung ein; der Fahrer zeigte keinerlei Reaktionen, um die Kollision zu vermeiden (National Highway Traffic & Safety Administration (NHTSA), 2017).

Am 18. März 2018 kollidierte ein modifizierter Volvo XC90, der zu Forschungszwecken von Uber Technologies, Inc. betrieben wurde, mit einer Fußgängerin in Tempe (Arizona) auf einer Landstraße bei ca. 63 km/h und verletzte sie tödlich. Die Fußgängerin querte die Straßenseite und schob dabei ein Fahrrad. Die Fahrerin („operator“) des Volvo wurde nicht verletzt. Im vorläufigen Bericht des National Transportation Safety Board (NTSB) (2018a) heißt es:

„The inward-facing video shows the vehicle operator glancing down toward the center of the vehicle several times before the crash. In a postcrash interview with NTSB investigators, the vehicle operator stated that she had been monitoring the self-driving system interface.“
(NTSB, 2018a, S. 3)

Am 23. März 2018 fuhr ein Tesla Model X mit aktiviertem Level 2 auf einem Highway in Mountain View (Kalifornien), auf dem zweiten Fahrstreifen von links. Als sich die Fahrbahn aufteilte, lenkte der Tesla nach links und kollidierte bei einer Geschwindigkeit von ca. 114 km/h mit einem (bereits beschädigten) Fahrzeugrückhaltesystem. Die vorherige Fahrt dauerte etwa 32 Minuten, währenddessen wurde die Level 2-Funktion viermal aktiviert. Unmittelbar vor der

Kollision fuhr der Fahrer 18 Minuten und 55 Sekunden mit aktiviertem Level 2. 15 Sekunden vor der Kollision erhielt der Fahrer eine akustische und zwei visuelle Warnungen, die Hände ans Lenkrad zu nehmen. In den 60 Sekunden vor der Kollision waren insgesamt für 34 Sekunden die Hände des Fahrers am Lenkrad. Die letzten sechs Sekunden vor der Kollision wurde keine Hand am Lenkrad detektiert (NTSB, 2018). In einem späteren Bericht der NTSB (2020) wurde zusätzlich Folgendes notiert:

„The Tesla driver was likely distracted by a gaming application on his cell phone before the crash, which prevented him from recognizing that Autopilot [(Level 2-Funktion)] had steered the SUV into a gore area of the highway not intended for vehicle travel.“ (NTSB, 2020, S. 2)

Weiter heißt es in dem Bericht:

„The Tesla Autopilot system did not provide an effective means of monitoring the driver’s level of engagement with the driving task, and the timing of alerts and warnings was insufficient to elicit the driver’s response to prevent the crash or mitigate its severity. Requirements are needed for driver monitoring systems for advanced driver assistance systems that provide partial driving automation (SAE Level 2 systems), and Tesla needs to develop applications that more effectively sense the driver’s level of engagement and that alert drivers who are not engaged.“ (NTSB, 2020, S. 2)

Im zuletzt zitierten Absatz wird bereits die Forderung nach Fahrerzustandserfassung deutlich. Um dies einordnen zu können, empfiehlt es sich, den Fahrerzustand beim automatisierten Fahren systematisch zu betrachten – bspw. mit Hilfe nachfolgenden Modells.

2.4.3 Fahrerzustand am Modell

Für ein grundsätzliches Verständnis der Fahrerzustandsproblematik beim automatisierten Fahren soll im Folgenden das Fahrerzustandsmodell von Marberger et al. (2018) vorgestellt werden (Abbildung 11), anhand dessen die in Abschnitt 2.4.5 präsentierte Forschung eingeordnet werden kann.

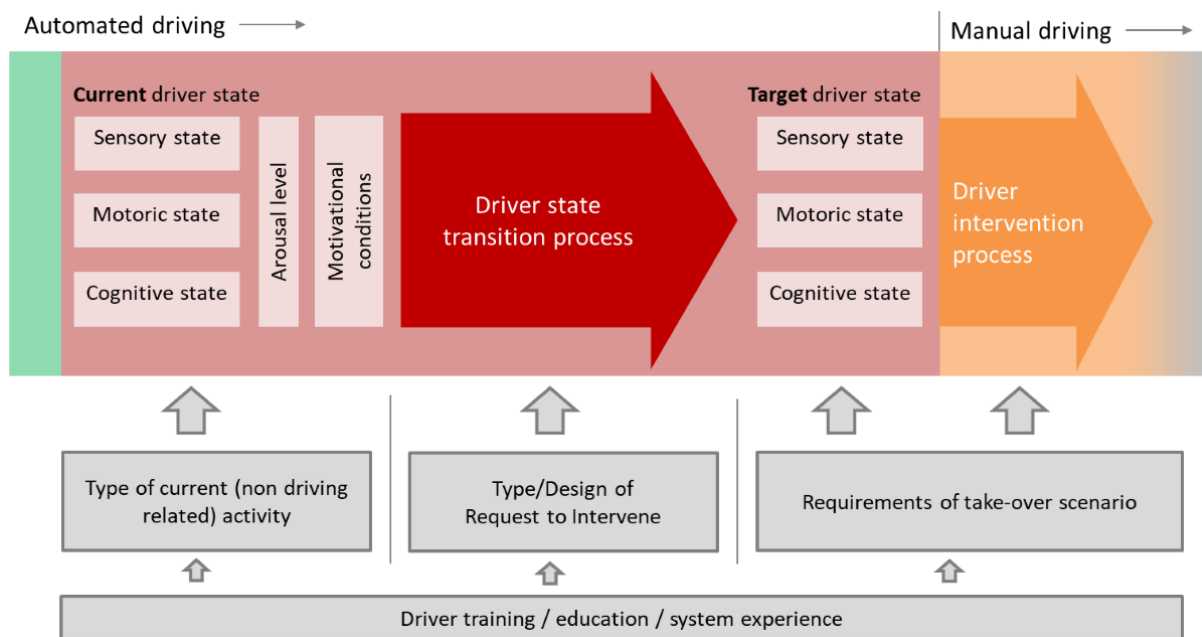


Abbildung 11: Fahrer Verfügbarkeitsmodell aus Marberger et al. (2018), S. 600.⁵

Die Autoren unterscheiden zunächst die Phase des automatisierten Fahrens und die sich daran anschließende Phase des manuellen Fahrbetriebs. Während die Automatisierung aktiv ist, besitzt der Fahrer einen spezifischen Zustand („current driver state“), der sich aus verschiedenen Modalitäten (sensorisch, motorisch und kognitiv) zusammensetzt. Er befindet sich ferner in einem bestimmten Arousal-Level und ist motiviert (intrinsisch wie extrinsisch), den gegenwärtigen Zustand einzunehmen bzw. beizubehalten. Maßgeblich wird dieser Zustand durch die aktuell auszuführende Tätigkeit moderiert (z.B. Überwachen der Fahrfunktion in Level 2 oder die Beschäftigung mit fahrfremden Tätigkeiten in Level 3). Bei Übernahme der manuellen Fahrzeugsteuerung wäre wiederum ein konkreter, ggf. anders gearteter Fahrerzustand optimal, der das anschließende manuelle Fahren verkehrssicher gewährleistet. Der „target driver state“ setzt sich ebenfalls aus den Modalitäten „sensorisch“, „motorisch“ und „kognitiv“ zusammen. Ein optimaler Zielzustand könnte beispielsweise folgende Attribute beinhalten: Ein umfassendes Situationsbewusstsein, eine maximal mögliche Wachheit sowie eine auf die Fahraufgabe ausgerichtete Aufmerksamkeit. Ferner könnten beide Hände das Lenkrad umgreifen und sich die Füße an der Pedalerie befinden. Wie das Modell beschreibt, gleicht nicht jeder Zielzustand dem anderen, sondern jedwedes (Verkehrs-)Szenario besitzt seine eigenen, spezifischen Anforderungen für den optimalen Übergang vom automatisierten zum manuellen Fahren. Zwischen ak-

⁵ Abgedruckt mit Genehmigung von Springer Nature: Springer, Cham, aus „Understanding and Applying the Concept of “Driver Availability” in Automated Driving“ von Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B., Copyright 2018.
https://link.springer.com/chapter/10.1007/978-3-319-60441-1_58

tuellem Fahrerzustand und Zielzustand kann ein Anpassungsprozess stattfinden, die sog. Übernahmephase, der „transition process“. In dieser Phase kann die Chance genutzt werden, den aktuellen Fahrerzustand möglichst nah an den Zielzustand heranzuführen. Je nach Automatisierungslevel gibt es beispielsweise unterschiedliche Zeitvorgaben, die das Erreichen des Zielzustands erschweren, aber auch erleichtern können. Darüber hinaus bestimmt die Ausgestaltung der Übernahmeaufforderung, das MMI-Design des „Request to Intervene“, maßgeblich, wie gut oder wie schnell der Fahrer die Aufforderung wahrnimmt und hierauf reagieren kann.

Aus dem Modell lassen sich verschiedene Implikationen für Stellgrößen ableiten: Je größer die Kongruenz zwischen aktuellem Fahrerzustand und Zielzustand, desto weniger Energie muss während des Übernahmeprozesses aufgewendet werden. Dies kann entweder erreicht werden, indem die Qualität des aktuellen Fahrerzustands gesteigert wird, oder indem die Anforderungen an den Zielzustand verringert werden. Demnach wäre insbesondere eine Kombination beider Maßnahmen sinnvoll. Der gesamte Prozess kann durch Expertise des Fahrers erleichtert und damit sicherer gestaltet werden (Marberger et al., 2018).

2.4.4 Experimentelle Methoden zur Erforschung des Fahrerzustands

Wie kann der Fahrerzustand beim automatisierten Fahren möglichst valide erfasst werden? Bereits in Abschnitt 2.3.5 wurden Müdigkeitsindikatoren vorgestellt, die üblicherweise beim manuellen Fahren Anwendung finden. Lediglich den Fokus auf abhängige Variablen bei der Konzeption einer experimentellen Studie zu richten, wäre zu kurz gegriffen. Stattdessen ist eine ganzheitliche Betrachtung der gewählten Forschungsmethoden vonnöten. In einer aktuellen Veröffentlichung haben Schömig, Befelein, Wiedemann und Neukum (2020) strukturiert sog. „Module“ für die Studienkonzeption einer Level 3-Untersuchung zusammengestellt. Grundsätzlich sollten sich viele der hier beschriebenen Ansätze auch auf Level 2-Untersuchungen übertragen lassen. Abbildung 12 gibt einen Überblick über die von Schömig et al. (2020) vorgeschlagene Modulaufteilung. Welche Modulanpassungen konkret für die Erforschung des Fahrerzustands sinnvoll sind, soll nachfolgend mit Hilfe der in Radlmayr et al. (2018) zusammengetragenen Expertise thematisiert werden.

Modul		Zu treffende Entscheidungen
01: Stichprobe		Wie viele und welche Probanden (bzgl. Alter, Geschlecht, Fahrerfahrung, Vorerfahrung mit Assistenz/automatisierten Systemen, weitere Auswahlkriterien)?
02: Simulatorenausstattung		Welche technischen Voraussetzungen muss der Simulator erfüllen?
03: Instruktion und Eingewöhnung		Wie müssen die Probanden instruiert werden? Welches Vorwissen sollen sie haben, wenn sie mit dem System interagieren?
04: Streckencharakteristika		Auf welchem Streckentyp soll die Studie stattfinden?
05: Systemgestaltung		Welche Eigenschaften soll das automatisierte System haben?
06: Mensch-Maschine-Schnittstelle		Welches Standard-HMI (engl. Human-Machine-Interface, kurz HMI) soll in der Studie verwendet werden?
07: Fahrerzustand		In welchem Zustand soll ein Fahrer für diese Art von Studie sein?
08: Fahrfremde Tätigkeiten		Soll sich der Fahrer mit nicht-fahrtbezogenen Tätigkeiten beschäftigen und wenn ja, mit welchen?
09: Prüf-szenario	Für Studien zur Übernahme-situation	Wie soll das Übernahmeszenario gestaltet sein, anhand dessen man die Übernahmefähigkeit eines Fahrers beurteilt?
	Für Studien zum Erleben des aktiven Systems	Welche Szenarien müssen berücksichtigt werden? Wie lange soll der Fahrer das aktive System erleben?
10: Abhängige Variablen	Als Maße der Übernahmeleistung	Welche Maße sollen für die Bewertung der Übernahmeleistung erfasst werden?
	Sonstige Maße der Systembewertung	Welche Maße können zur Bewertung des Systems bzw. der Mensch-Maschine-Schnittstelle genutzt werden?

Abbildung 12: Zu treffende methodische Entscheidungen für die Konzeption von Level 3-Untersuchungen (aus Schömig et al., 2020, S. 8).

Die Frage nach einer adäquaten Stichprobengröße und -zusammensetzung stellt sich durchweg vor jeder Studienplanung. Allerdings geben die im Vorfeld zu erwartenden individuellen Unterschiede im Ermüdungsverhalten Grund zum Anlass, besondere Aspekte zu berücksichtigen. Einen sensiblen Bereich nimmt daher die Rolle der schlafbezogenen Müdigkeit im Fahrversuch ein (vgl. Abschnitt 2.3.3). Es kann hilfreich sein, Probanden zu vergleichbaren Tageszeiten zu testen und ihre Schlafqualität sowie ihren Koffeinkonsum aufzuzeichnen. Asynchron zum Chronotyp durchgeführte Untersuchungen können Effektstärken erhöhen: Üblicherweise wird der Stichprobenumfang in experimentalpsychologischen Arbeiten auf Basis der zu erwartenden Effektstärke abgeschätzt. In ingenieurpsychologischer Anwendungsforschung ist nicht selten eben diese Effektstärke Gegenstand der Untersuchung. Bei unbekannter Effektgröße, wie es bei Müdigkeitsmessung im automatisierten Fahrkontext gegenwärtig noch der Fall ist (vgl. Abschnitt 2.4.1), könnte zumindest im Nachhinein die Sensitivität der verwendeten Müdigkeitsindikatoren bei gegebener Stichprobengröße diskutiert werden. Für inferenzstatistische Analysen ist ein hoher Stichprobenumfang komfortabel – allerdings steigen mit zeitlicher Dauer und methodischem Aufwand die Kosten der Untersuchung. Daher müssen ökonomische Restriktionen

beachtet werden, sodass hinsichtlich weniger Merkmale stratifizierte, jedoch kleinere Stichproben empfehlenswert sind (vgl. hierzu Vollrath, 2015). Insgesamt sollte die Erforschung des Fahrerzustands nach Möglichkeit repräsentativ erfolgen, d.h. sich zumindest auf jeden Fahrer jedes Alters und Geschlechts übertragen lassen. Dies wäre auch im Hinblick auf verschiedene Vorerfahrungen mit automatisierten Fahrfunktionen wünschenswert (vgl. Modul 1, Abbildung 12).

Unter Modul 2 und 4 subsummieren Schömig et al. (2020) die Simulatorenausstattung und die zu programmierenden Versuchsstrecken. Damit wird impliziert – und letztlich auch empfohlen – einen Fahrsimulator für eine Level 3-Untersuchung zu verwenden. Leider werden von den Autoren alternative Möglichkeiten, wie eine Teststreckenstudie, eine Fahrt im realen Straßenverkehr oder (unabhängig von der Fahrumgebung) die Zuhilfenahme eines sog. (nachfolgend erläuterten) „Wizard-of-Oz-Fahrzeugs“ ausgespart. Hier geht es um die klassische Abwägung, ob eine höhere Kontrollierbarkeit (hohe interne Validität) oder eine höhere Übertragbarkeit auf reale Anwendungen (hohe externe Validität) angebracht sind.

Ermüdungsmessung im Fahrsimulator kann die Problematik mit sich bringen, unerwünschte schlafbezogene Einflussfaktoren zu triggern. Beispielsweise sollten Effekte durch Verdunklung, niedrige Temperaturen (Schmidt, 2020) und/oder auftretende „Simulator-Sickness“ nicht außer Acht gelassen werden. So konnten Fors, Ahlstrom und Anund (2018) beim manuellen Fahren zufallskritisch nachweisen, dass Probanden im Fahrsimulator eine höhere subjektive Müdigkeit (über KSS ermittelt), einen verlängerten Lidschluss, weniger Blicke auf die Straßemitte und eine höhere Fahrgeschwindigkeit aufwiesen als dies bei vergleichbarer Testung im realen Straßenverkehr der Fall war. Hingegen kann der reale Straßenverkehr vielerlei unerwartete (Verkehrs-) Situationen hervorbringen, selbst wenn eine konstante Route gefahren wird (Klamroth et al., 2019). Mögliche Störeinflüsse, die auf den Probanden kurzzeitig aktivierend wirken (z.B. notwendige Fahrstreifenwechsel, schnell überholende Fahrzeuge oder Einscherer vor das eigene Fahrzeug) sind zwar realitätsnah, erschweren jedoch die Validitätsbestimmung eines Müdigkeitsindikators. Eine monotone Teststrecke kann hierfür vorteilhafter sein – sobald man sich jedoch hierfür entscheidet, wird auch ein Versuchsträger benötigt, der über automatisierte Fahrfunktionen verfügt und eine entsprechende Messdatenaufzeichnung leisten kann. Im Fahrsimulator kann die Parametrisierung der Funktion entsprechend programmiert werden, wobei im realen Kontext Gas-, Brems- und Lenkroboter hierfür das Pendant bilden würden. Diese können jedoch bei höheren Fahrgeschwindigkeiten schnell an ihre Reliabilitäts- und Sicher-

heitsgrenzen stoßen. Aus diesem Grund erlangt der Einsatz eines sog. „Wizard-of-Oz-Fahrzeugs“ wachsende Beliebtheit. Auch in vorliegender Arbeit spielt dieser methodische Ansatz eine wesentliche Grundlage für die Versuchsdurchführung und wird nun näher betrachtet.

Die Idee, die einem Wizard-of-Oz-Fahrzeug zugrunde liegt, ist nicht neu. Die Namensgebung geht dabei auf die Erzählung „The Wonderful Wizard of Oz“ des US-amerikanischen Schriftstellers Lyman Frank Baum aus dem Jahr 1900 zurück, in der die Figur „Der Zauberer von Oz“ mit Hilfe von Illusionen und Effekten seine vermeintlichen Zauberkünste gegenüber den Protagonisten lediglich vortäuscht (Baum, 1900). Ebenso verhält es sich im übertragenen Sinn auch beim Wizard-of-Oz Prinzip in der experimentalpsychologischen Forschung: Erste Anwendungen außerhalb des Automotive-Bereichs gab es in den 70er und 80er Jahren (vgl. für eine Übersicht Bengler, Omozik & Müller, 2019): Sobald die Wirkung technischer Systeme auf den Menschen evaluiert werden soll, jedoch die technologischen Möglichkeiten für eine (ökonomisch darstellbare) Umsetzung (noch) nicht gegeben sind, kann auf die Wizard-of-Oz-Methodik zurückgegriffen werden. Ziel dabei ist es, Versuchsteilnehmern eine möglichst perfekte Nachbildung der Arbeitsweise eines technischen Systems vorzutäuschen – im besten Fall, ohne dass die Probanden über die Illusion in Kenntnis sind. Auf das automatisierte Fahren übertragen hieße das, die Regelungstätigkeit der automatisierten Fahrfunktion nachzuahmen, und zwar durch eine verdeckte Regelungstätigkeit eines menschlichen Fahrers („Wizard-Fahrer“). Vorteile ergeben sich dadurch, dass verschiedene Automatisierungslevel gemäß SAE (vgl. Abschnitt 2.1) in ein- und demselben realen Fahrzeug dargestellt werden können. Auf diese Weise lassen sich Merkmale hoher interner und externer Validität bei einem Fahrversuch im öffentlichen Straßenverkehr oder auf einer Teststrecke verknüpfen, wie bereits einige Studien demonstrierten (Baltodano, Sibi, Martelaro, Gowda & Ju, 2015; Berghöfer, Purucker, Naujoks, Wiedemann & Marberger, 2018; Hoeger et al., 2011; Manstetten, Marberger & Beruscha, 2019; Radlmayr et al., 2018). Aktuell fehlen jedoch noch standardisierte Gütekriterien, um die Ergebnisse aus Wizard-of-Oz-Versuchen sinnvoll vergleichen zu können (Müller, Weinbeer & Bengler, 2019). Hierfür haben Müller et al. (2019) die in Tabelle 4 dargestellten methodischen Herausforderungen identifiziert.

Tabelle 4: Methodische Herausforderungen bei Wizard-of-Oz-Fahrversuchen abhängig vom Testgütekriterium (nach Müller et al., 2019).

Testgütekriterium	Methodische Herausforderung
Objektivität	Verschiedene Wizard-Fahrer müssen untereinander in der Lage sein, den gleichen Fahrstil zu reproduzieren.
Reliabilität	Ein Wizard-Fahrer muss in der Lage sein, den gleichen Fahrstil zu verschiedenen Zeitpunkten zu reproduzieren.
Validität	Die Illusion des Wizard-of-Oz-Verfahrens muss permanent aufrechterhalten werden. Die simulierte automatisierte Fahrfunktion muss sich wie eine reale automatisierte Fahrfunktion verhalten.

Selbst wenn ein Wizard-of-Oz-Fahrzeug für die Untersuchung verwendet wird, muss entschieden werden, welche Leistungsgrenzen die (fiktive) Automatisierungsfunktion besitzen soll, bzw. wie zuverlässig die Funktion ist (Modul 5 nach Schömig et al., 2020). Aufforderungen zur Übernahme der Fahrzeugsteuerung sollten möglichst realitäts- d.h. seriennah angezeigt werden und vonstattengehen (Modul 6), wenn bei der Erforschung des Fahrerzustands das Ermüdungsverhalten bezogen auf das Automatisierungslevel im Vordergrund steht, und nicht der Fokus auf HMI-spezifischen Fragestellungen an sich gelegt wird.

Der Einsatz eines Wizard-of-Oz-Fahrzeugs kann je nach technischem Aufbau besondere Restriktionen in Bezug auf eine realitätsnahe Umsetzung eines SAE Levels bedingen. Die technische Umsetzung des in vorliegender Arbeit verwendeten Wizard-of-Oz-Fahrzeugs sieht beispielsweise vor, dass eine parallele Regelungstätigkeit der Längs- und/oder Querführung zwischen Proband und Wizard-Fahrer zu keinem Zeitpunkt stattfindet. Insofern muss der Proband auch während Level 2 permanent die Hände auf seinen Oberschenkeln ablegen, also „hands-off“ fahren. Eine umfassende Einordnung von Auswirkungen freihändiger Überwachung einer Level 2-Funktion findet sich in Josten, Zlocki und Eckstein (2016). Demnach sollte mit verlängerten Reaktionszeiten in Übernahme-situationen gerechnet werden, selbst wenn Josten et al. (2016) die zufallskritische Absicherung dieser Beobachtung auf einer Teststrecke verwehrt blieb.

Eine entscheidende Rolle spielt das Zusammenwirken von Modul 3 (Instruktion und Eingewöhnung) mit Modul 8 (Fahrerfremde Tätigkeiten) bei der methodischen Planung, wenn Ermüdungsverhalten sichtbar gemacht werden soll. Die Fahrt mit einem Level 2 ermöglicht keine FFT-Beschäftigung. Level 3 sieht dies prinzipiell vor, verpflichtend ist die Beschäftigung nicht. Daraus resultiert ein methodisches Dilemma: Entweder wird mit höherer externer Validität eine vergleichende Untersuchung zwischen Level 2 und Level 3 *mit* FFT durchgeführt, oder die Untersuchung wird zugunsten höherer interner Validität auf einen Vergleich zwischen Level 3 *mit und ohne* FFT beschränkt, um beispielsweise den Effekt von FFT auf das Ermüdungsverhalten ohne Konfundierung der Überwachungstätigkeit sichtbar zu machen. In jedem Fall müssen die Probanden zweifelsfrei über die konkrete Arbeitsteilung zwischen Automatisierungsfunktion und Fahreraufgabe instruiert werden, damit nicht bereits hierdurch eine Quelle für „Mode Confusion“ entsteht, und die Ergebnisinterpretation somit erschwert wird.

Modul 3 nach Schömig et al. (2020) ermahnt auch zur Eingewöhnung der Probanden vor der primären Versuchsfahrt. Insbesondere bei Müdigkeitserfassung besteht die Problematik, dass Probanden womöglich sehr motiviert in die Studie hineingehen und sehr wahrscheinlich eine Automatisierungsfunktion höheren Levels zum ersten Mal überhaupt in Aktion erleben. Dies kann aktivierend wirken, wodurch das tatsächliche Müdigkeitsniveau bei routinierter Nutzung unterschätzt wird. Die Module 9 und 10 müssen hingegen bei Fahrerzustandsuntersuchungen entsprechend angepasst werden. Schließlich soll das Konstrukt des Fahrerzustands (nach Schömig et al. (2020) Modul 7) im vorliegenden Fall als abhängige Variable dienen. Verschiedene Indikatoren sind hierfür denkbar und beanspruchen ihrerseits am validesten die Fahrermüdigkeit während der automatisierten Fahrt zu erfassen (Hecht et al., 2018). Dem Review von Hecht et al. (2018) zu Folge gibt es keine immensen Validitätsunterschiede der Indikatoren zwischen manuellem Fahren (vgl. Abschnitt 2.3.5) und ihrer Anwendung beim automatisierten Fahren. Eye-Tracking kann eine hohe Sensitivität und Spezifität besitzen wie auch EEG, mit den bekannten Praktikabilitätsproblemen. Feldhütter, Feierle, Kalb und Bengler (2018) berichten anhand ihres entwickelten objektiven Messverfahrens, das mit Fusion aus Lidschlusslevel und Kopfbewegung des Fahrers arbeitet, von einer Sensitivität von 90.0 % und einer Spezifität von 99.2 % (unter kontrollierten Bedingungen im Fahrsimulator gemessen). Für EEG-basierte Verfahren ist in dieser Hinsicht beim automatisierten Fahren noch keine Abschätzung erfolgt. Generell wird empfohlen, auch im automatisierten Kontext von subjektiven Selbstbewertungen der Müdigkeit nach Möglichkeit Abstand zu nehmen (Hecht et al., 2018).

Als verhaltensbasierter Müdigkeitsindikator kann beim automatisierten Fahren auch die Übernahme der manuellen Fahrzeugsteuerung nach einer längeren Level 2 oder Level 3-Periode dienen. Jarosch et al. (2019) erklären – in Anlehnung an das zuvor vorgestellte Fahrerverfügbarkeitsmodell nach Marberger et al. (2018), dass der Übernahmeprozess durch passive aufgabenbezogene Müdigkeit moduliert sein kann. Nach einer Level 3-Fahrt müssen insbesondere die Merkmale der zuvor bearbeiteten FFT berücksichtigt werden: Es sollte zwischen leicht und schwer unterbrechbaren FFT differenziert werden, wobei eine hohe Motivation einer vermeintlich leicht zu unterbrechenden FFT weiter nachzugehen, eine verzerrte Ableitung der tatsächlich vorliegenden Müdigkeit bedingen kann (Jarosch et al., 2019). Einige Untersuchungen bedienen sich daher einer standardisierten Tätigkeit wie der visuellen „Surrogate Reference Task“ (SuRT; ISO 14198, 2012) oder der kognitiven „n-back Task“ (Gold, 2016; Klamroth et al., 2019; Radlmayr, Gold, Lorenz, Farid & Bengler, 2014). Dabei zeigte sich, dass die SuRT während Level 3 zu längeren Übernahmereaktionen führt als eine „2-back Task“ (Gold, 2016). Zu bedenken ist insgesamt, dass diese Tätigkeiten aufgrund ihres marginal motivierenden Charakters lediglich eingeschränkt geeignet sind, um wiederum energetische Aktivierung der Fahrer durch die Bearbeitung von FFT sichtbar zu machen (vgl. Jarosch, Kuhnt, Paradies & Bengler, 2017; Abschnitt 2.4.5.2).

Bezogen auf das experimentelle Design eines Fahrversuchs, der müdigkeitsbedingte Effekte sichtbar machen soll, sind zwei fundamental unterschiedliche Ansätze vorstellbar. Beiden Ansätzen liegt zugrunde, dass voraussichtlich mit großen interindividuellen Unterschieden im Ermüdungsverhalten umgegangen werden muss. In der ersten Variante fahren die Versuchsteilnehmer so lange automatisiert, bis ein vorab definiertes Müdigkeitsniveau erreicht ist. Die Fahrdauer fungiert dann als abhängige Variable. Dieser Ansatz ist mit verschiedenen Vorteilen verbunden: Der Versuchsleiter kann sich (relativ) sicher sein, dass der jeweilige Proband nach einer bestimmten Zeit einen anvisierten Zustand besitzt. Aus diesem Fahrerzustand heraus kann dann beispielsweise eine Übernahmeaufforderung zum manuellen Fahren erfolgen, um die Performanz des Versuchsteilnehmers zu messen (Kreuzmair, Gold & Meyer, 2017; Schömig & Kaussner, 2010; Weinbeer et al., 2017). Dieses Vorgehen kann sehr sinnvoll sein, wenn a) bekannt ist, dass der Müdigkeitsindikator valide genug ist, auch die Änderungen im Fahrerzustand zu erfassen und b) bekannt ist, wie lange in etwa die Fahrt im gegebenen Szenario andauern muss, um die gewünschten Fahrerzustandsänderungen zu erreichen. Andernfalls dauern die Fahrten womöglich sehr lange oder Probanden erreichen nie ein bestimmtes Müdigkeitsniveau (siehe z.B. Weinbeer et al., 2017). In der zweiten Variante wird im Versuchsdesign eine konkrete Fahrdauer festgelegt. Dieser Ansatz besitzt einen höheren explorativen Charakter als die erste Variante. Grundsätzlich kann sich auch hier der Versuchsleiter nicht sicher sein, dass alle

Probanden im gewünschten Maße ermüden oder dass eine Intervention zu einem bestimmten Fahrerzustandsmoment stattfindet. Jedoch besitzt diese Variante den Vorteil, versuchsökonomisch restriktiv zu sein – das macht vor allem Sinn, wenn die Validität des Müdigkeitsindikators unklar ist. Trotzdem ist es möglich zu bestimmen, welches Müdigkeitsniveau zu welchem Zeitpunkt der Fahrt vorlag (bei gegebener Validität), und Effekte durch Übernahmeaufforderungen sowohl zeit- als auch müdigkeitsabhängig zu beurteilen. Die meisten Untersuchungen (vgl. Tabelle 5 im nächsten Abschnitt 2.4.5) bedienen sich dieser Variante.

2.4.5 Ermüdung und Vigilanzdekrement: Studienlage

Nachfolgend soll auf bisherige Forschung eingegangen werden, die zum Ziel hatte, Müdigkeit oder Vigilanzdekrement (als Konsequenz von Ermüdung) in Bezug auf den Fahrerzustand beim automatisierten Fahren zu erfassen. Die mehrseitige Tabelle 5 gibt eine Übersicht über relevante Studien (alphabetisch geordnet nach Autoren), die die Erforschung von Ermüdung des Fahrers und seinem resultierenden Vigilanzdekrement zum Gegenstand hatten. Die Übersicht differenziert zwischen einzelnen SAE Leveln; häufig dient manuelles Autofahren (Level 0) als Referenz. Es wurde entschieden, ebenfalls Studien mit aufzunehmen, die zeitlich bedingte Leistungseinbußen des Fahrers bei Verwendung eines Level 1 untersuchten (hier ausschließlich die Automatisierung der Längsführung betreffend). Schließlich obliegt es dem Fahrer bei Verwendung eines Level 1 – analog zu einem Level 2 – die Regelungstätigkeit der Funktion und das Verkehrsumfeld permanent zu überwachen. Folglich muss der Fahrer auch hierbei eine Vigilanzaufgabe ausführen. Diese tritt allerdings nicht in „Reinform“ wie bei Nutzung eines Level 2 in Erscheinung, da der Fahrer immer noch eigene Regelungstätigkeiten leistet (hier die Querführung betreffend). Ferner zeigte die Literaturrecherche, dass es verschiedene Untersuchungsansätze für Level 3-bedingte Effekte gibt. So muss in Tabelle 5 zwischen Müdigkeitseffekten unterschieden werden, die während einer Level 3-Fahrt *ohne* die fortwährende Bearbeitung fahrfremder Tätigkeiten (FFT) identifiziert wurden, und solchen, die *mit* Bearbeitung von FFT berichtet wurden. Häufig fehlt es leider an einer präzisen Beschreibung der Instruktion gegenüber den Probanden und den ihnen erläuterten Leistungsgrenzen der Automatisierungsfunktion: Wenn sowohl Längs- als auch Querführung automatisiert wurden und dem Fahrer keine FFT instruiert wurde, besteht die Frage, was er zu tun hatte. Sollte er lediglich „die Fahrt genießen“ (siehe z.B. Vogelpohl et al., 2017) oder ist eine Überwachungstätigkeit vonnöten (siehe z.B. Schömig, Hargutt, Neukum, Petermann-Stock & Othersen, 2015). Nicht nur die Einordnung in das entsprechende SAE-Level fällt damit schwerer, sondern die Anforderungen

an den Fahrer und der (motivationalen) Aufrechterhaltung seines Zustands ist hiervon maßgeblich beeinflusst. Dennoch bestand hier der Anspruch, eine möglichst eindeutige Zuordnung zu einem SAE Level auf Basis des Forschungsziels der jeweiligen Studie zu gewährleisten.

In der Tabellenspalte „Untersuchungsszenario“ wird aufgelistet, welche Prüfumgebung (Fahr-simulator oder Realfahrzeug) der Studie zugrunde lag. Das Augenmerk wurde konkret auf den Monotoniegrad gerichtet, der über die jeweils berichtete Verkehrsdichte und Textur der Szene-rie abgeschätzt wurde. Die vorherrschende Fahrgeschwindigkeit bietet hierzu eine ergänzende Information. In der Spalte „Müdigkeits- oder Vigilanzindikator“ wird auf die im Fahrversuch verwendeten abhängigen Variablen fokussiert, wobei die Auswirkung der experimentellen Ma-nipulation in der Tabellenspalte rechts daneben erläutert wird. Einige der hier kompakt zusam-mengefassten Studienergebnisse werden im Folgenden näher diskutiert. Dabei sollen Unter-schiede aber auch Gemeinsamkeiten – im Speziellen die Entwicklung von Müdigkeit und ihre Einflussfaktoren betreffend – evaluiert werden.

Tabelle 5: Übersicht über experimentelle Studien mit dem Fokus auf „Müdigkeit und/oder Vigilanzdekrement beim automatisierten Fahren“.

Autoren	Stichprobe	SAE Level	Untersuchungsszenario	Fahrdauer	Müdigkeits- oder Vigilanzindikator	Auswirkung
Biondi et al. (2018)	$N = 22$	0 und 2	Realfahrzeug, Realverkehr; geringe Verkehrsdichte, Autobahn, ca. 120 km/h	je Level ca. 50 min	EKG (Herzrate, HRV), TUT, RZ (auf vibrotaktile Stimulation)	sig. längere RZ und niedrigere Herzrate bei L2 vs. L0 über gesamte Fahrdauer; keine sig. Unterschiede für HRV und TUT.
Feldhütter, Gold, Schneider und Bengler (2017)	$N = 31$	3 (ohne und mit FFT)	Fahrsimulator, statisch, vermutlich geringe Verkehrsdichte, Autobahn, 120 km/h	je Level 5 min vs. 20 min	RZ (auf TOR); Blickdauer auf Fahrszenerie (ohne FFT)	sig. längere RZ mit FFT (SuRT) vs. ohne FFT; sig. schnellere RZ nach 5 min vs. 20 min; In 20-minütiger Bedingung: sig. weniger (Kontroll-)Blicke auf Szenerie nach 18 min vs. 3 min.
Gonçalves, Happee und Bengler (2016)	$N = 31$	2	Fahrsimulator, statisch, keine anderen Verkehrsteilnehmer, hohe Monotonie der Umgebung	3 min (wach) vs. ca. 17 min (müde)	SSS, Lidschlusslänge + Häufigkeit, PERCLOS, Fahrverhalten nach TOR	hohes Müdigkeitslevel (Stufe 5 von 6 bei SSS) nach im Mittel 17 min.; Validere Messung durch SSS als PERCLOS möglich; sig. höhere Querschleunigung bei Müden als bei Wachen.

Jarosch et al. (2017)	$N = 56$	3 (mit FFT)	Fahrsimulator, dynamisch; geringe bis moderate Verkehrsdichte, Autobahn, 120 km/h	2x 25 min	KSS, PERCLOS, Lidschlusslänge- und Häufigkeit	kein sig. Müdigkeitsanstieg (subjektiv wie objektiv) beim Quiz; sig. Müdigkeitsanstieg (subjektiv wie objektiv) ab 10 min bei monotoner FFT, monoton steigend; nach TOR sig. Müdigkeitsreduktion.
Jarosch, Paradies, Feiner und Bengler (2019a)	$N = 42$	3 (mit FFT)	Wizard-of-Oz, Realverkehr, moderate Verkehrsdichte, Autobahn, 110 km/h	2x 30 min	KSS, PERCLOS	kein sig. Müdigkeitsanstieg (subjektiv wie objektiv) bei frei wählbarer FFT; sig. Müdigkeitsanstieg (subjektiv wie objektiv) ab 15 min bei monotoner FFT, monoton steigend.
Körber et al. (2015)	$N = 20$	2	Fahrsimulator, statisch, normale Verkehrsdichte, Autobahn, 125 km/h	42,5 min	Lidschlusslänge + Häufigkeit + Pupillendurchmesser, PERCLOS, TUT, RZ (auf auditives Oddball Paradigma)	keine sig. RZ-Veränderung mit der Fahrdauer; sig. Anstieg der Lidschlussdauer mit der Fahrdauer (bis ca. Minute 15); sig. Reduktion des Pupillendurchmessers (bis ca. Minute 25); sig. Zunahme der Lidschlüsse ca. ab Minute 35); kein sig. Effekt anhand von PERCLOS bestimmbar; sig. Zunahme von TUT zwischen Anfang und Ende der Fahrt.

Kreuzmair et al. (2017)	$N = 22$	3 (ohne FFT)	Fahrsimulator, dynamisch; hohe Monotonie, Autobahn, Nachtfahrt, 120 km/h	keine Angabe, müdigkeitsabhängig	KSS (Eigen- und Fremdbewertung) sowie PERCLOS (als Kriterium für Fahrtdauer), Übernahmequalität als AV	Keine nennenswerten müdigkeitsabhängigen Effekte bezogen auf die Übernahmequalität, da sich (entsprechend der Autoren) Lerneffekte bei sich wiederholender TOR mit etwaigen Müdigkeitseffekten überlagern.
Kundinger, Riener, Sofra und Weigl (2018)	$N = 30$	0 und 2	Fahrsimulator, dynamisch; geringe Verkehrsdichte, monotone Autobahn, 110 km/h	je Level 45 min	KSS (Eigenbewertung während und nach der Fahrt)	sig. Müdigkeitsanstieg (unabhängig vom Level) über die Fahrtdauer; sig. höherer Müdigkeitsanstieg bei L2 vs. L0; sig. höhere Müdigkeitslevel bei Jüngeren (20-25 Jahre) vs. Älteren (65-70 Jahre).
Kundinger, Riener, Sofra und Weigl (2020)	$N = 30$	0 und 2	Realfahrzeug, Teststrecke (Fahrroboter); sehr hohe Monotonie, 25 km/h	je Level 45 min	KSS (Eigenbewertung während und nach der Fahrt), Herzrate	sig. Müdigkeitsanstieg (unabhängig vom SAE Level) über die Fahrtdauer; sig. höherer Müdigkeitsanstieg bei L2 vs. L0; sig. höhere Müdigkeitslevel bei Jüngeren (20-25 Jahre) vs. Älteren (65-70 Jahre); niedrigere Müdigkeitslevel am Abend; inkonsistente Korrelationen zwischen KSS und Herzrate; Kontinuierlich niedrigere Herzrate bei L2 vs. L0, keine Veränderung über die Fahrtdauer.

Miller et al. (2015)	$N = 48$	2 und 3 (mit FFT)	Fahrsimulator, (Typ unbekannt); sehr geringe Verkehrsichte, monotone Landstraße, vermutlich zwischen 80 und 100 km/h	je Level 8,5 min, jeweils 3 mal (3 Blöcke)	Videoanalyse: Gähnen + Lid-schluss > 5s	In Block 2 und 3: sig. höhere Müdigkeit bei L2 vs. L3; sig. schnellere RZ auf TOR bei L2 vs. L3 nach Block 3.
Naujoks, Höfling, Purucker und Zeeb (2018)	$N = 64$	2 und 3 (mit FFT)	Fahrsimulator, dynamisch; moderate Verkehrsdichte, Autobahn, 130 km/h	je Level 50 min vs. 110 min	Videoanalyse, Workload, RZ auf kritisches Ereignis nach Übernahme	Müdigkeit, Workload und Motivation waren sig. Prädiktoren für RZ bei L2, nicht für L3. Keine weiteren nennenswerten Unterschiede zwischen L2 und L3.
Neubauer et al. (2014)	$N = 180$	0, 1, 3 (ohne FFT und mit FFT)	Fahrsimulator, statisch, keine konkreten Angaben zum Verkehrsszenario	je Level 45 min	Workload, RZ auf kritisches Ereignis	sig. höherer Workload für L0 vs. L3 (ohne FFT); sig. längere Bremsreaktion auf Ereignis nach 42 min für L1 vs. L0, sowie L3 (mit FFT) vs. L0.

Saxby, Matthews, Warm, Hitchcock und Neubauer (2013), Studie 2	$N = 168$	0 und 2	Fahrsimulator, statisch, kein Verkehr im Egofahrstreifen; hoher Detailgrad der Szenerie, urban, vermutlich zwischen 50 und 60 km/h	je Level 10 min vs. 30 min	Stress (Aufgaben-involviertheit), SDLP über L0 im Anschluss, RZ (auf kritisches Ereignis)	sig. niedrigerer Stress in L2 vs. L0, sig. Reduktion mit erhöhter Fahrdauer, stärkerer Effekt für L2; sig. niedrigere SDLP nach L2 vs. L0; sig. erhöhte RZ (Bremsen und Lenken) nach L2 vs. L0; sig. niedrigere Kollisionsvermeidung in L2 vs. L0, nach 30 min werden weniger Kollisionen vermieden als nach 10 min; L2 für 30 min: alle Probanden kollidierten.
Schömig und Kaussner (2010)	$N = 6$	2	Fahrsimulator, dynamisch, normale Verkehrsdichte und Stau, Autobahn, 110 km/h	müdigkeitsabhängig	Kopf- und Augenbewegungen	Erfolgreiche Validierung der Müdigkeitsdetektion; inkonsistente Erholungseffekte auf TOR.
Schömig et al. (2015)	$N = 16$	0, 3 (ohne FFT und mit FFT)	Fahrsimulator, dynamisch, geringe Verkehrsdichte, Autobahn, 120 km/h	viermal 15 min, Start einer 15-minütigen Sequenz müdigkeitsabhängig	Lidschlusslänge-Häufigkeit + Lidöffnungslevel	L0 und L3 (ohne FFT): größter Müdigkeitsanstieg in den ersten 15 Minuten; zwischen drittem und viertem 15-Minuten-Block keine Müdigkeitsveränderung; größter Müdigkeitsanstieg für L3 (ohne FFT); keine Müdigkeitsänderung bei L3 (mit FFT); L3 (ohne FFT): nach TOR deutlicher Erholungseffekt.

Stapel, Mullakkal- Babu und Happee (2017)	$N = 16$	0 und 2	Realfahrzeug, Realverkehr; geringe vs. normale Ver- kehrsdichte, Autobahn, ca. 90 km/h	je Level und Ver- kehrs- dichte ca. 30 min	NASA-TLX, au- ditive Detection Response Task (DRT)	sig. geringerer Workload bei L2 vs. L0 (nur für erfahrene L2-Nutzer); sig. geringerer Workload in monotoner Umgebung (unabhängig ob L0 oder L2); sig. schnellere RZ auf DRT in monoto- ner Umgebung bei erfahrenen L2-Nut- zern; sig. schnellere RZ auf DRT in L0 vs. L2 bei normaler Verkehrsdichte bei uner- fahrenen L2-Nutzern; sig. mehr Auslassungen bei normaler vs. monotoner Verkehrssituation (unabhän- gig ob L0 oder L2); sig. mehr Auslassungen bei unerfahrenen L2-Nutzern in L2 vs. L0 (bei normaler Verkehrsdichte).
Vogelpohl et al. (2017)	$N = 60$	0 und 3 (ohne FFT)	Fahrsimulator, statisch, mo- derate Ver- kehrsdichte, Dunkelheit, 120 km/h	je Level 55 min	KSS, Videoana- lyse, RZ (ver- schiedene Indi- katoren auf TOR (nach 5s))	L3: besonders starker Müdigkeitsanstieg nach 15-35 Minuten; L0: besonders starker Müdigkeitsanstieg nach 45-50 Minuten; sig. spätere Ermüdung ohne Schlafdepri- vation, sig. frühere Ermüdung mit Schlafdeprivation; späterer Blick auf Tacho bei L3 vs. L0, mehr Bremsreaktionen auf bremsendes Vorderfahrzeug bei L3 vs. L0, umge- kehrt mehr Fahrstreifenwechsel ohne Bremsen bei L0 vs. L3.

Vollrath, Schleicher und Gelau (2011)	$N = 22$	0 und 1	Fahrsimulator, dynamisch, ge- ringe Verkehrs- dichte, Auto- bahn	je Level ca. 60 min	Workload, sub- jektives Sicher- heitsgefühl, RZ bei Geschwindig- keitsanpassung	sig. geringerer Workload und höheres Sicherheitsgefühl bei L1 vs. L0; sig. verlängerte Geschwindigkeitsanpas- sungsdauer bei L1 vs. L0 (Differenz ca. 5s).
Weinbeer et al. (2017)	$N = 31$	3 (ohne FFT)	Wizard-of-Oz, Realverkehr, normale Ver- kehrsdichte, Autobahn, 130 km/h	müdig- keitsab- hängig	Beobachterrating (angelehnt an Wierwille und Ellsworth (1994), um eine Stufe er- weitert), RZ auf TOR	nach 15 Minuten: 20 % der Probanden ha- ben ein mittleres, 0 % ein hohes Müdig- keitslevel erreicht; nach 30 Minuten: ca. 47 % der Probanden haben ein mittleres, ca. 17 % ein hohes Müdigkeitslevel erreicht; nach 60 Minuten: ca. 73 % der Probanden haben ein mittleres, ca. 57 % ein hohes Müdigkeitslevel erreicht; kein sig. Müdigkeitseffekt auf TOR.

Anhand Tabelle 5 wird deutlich, dass das Thema Ermüdung und Vigilanzdekrement beim automatisierten Fahren mit verschiedenen experimentellen Methoden innerhalb des letzten Jahrzehnts (mit den meisten Veröffentlichungen innerhalb der letzten fünf Jahre) beforscht wurde (vgl. auch Abschnitt 2.4; Abbildung 10). Dabei ist das gesamte Spektrum an SAE Leveln (bis Level 3) abgedeckt – jedoch mit unterschiedlichen Paarvergleichen, um teils spezifische Fragestellungen zu beantworten. Nicht zwangsläufig wurde eine manuelle Fahrreferenz verwendet, einige Studien betrachten Fahrerzustandseffekte stattdessen isoliert für das jeweilige SAE Level. Mitunter werden auf diese Weise Validitätsvergleiche zwischen verschiedenen Müdigkeits- und/oder Vigilanzindikatoren ermöglicht. Auffällig ist demnach die Heterogenität der zugrunde gelegten Indikatoren, aber auch die jeweilige automatisierte Fahrdauer – eine Einordnung bisheriger Erkenntnisse im Bereich der Fahrerzustandserforschung profitiert von der Vielfalt der experimentellen Untersuchungsansätze.

2.4.5.1 Studienlage für SAE Level 2

Bereits bei Verwendung eines *Adaptive Cruise Control* (ACC; SAE Level 1) muss der Fahrer Überwachungsaufgaben ausführen. Vollrath et al. (2011) berichten hier vor einer eklatant verzögerten Reaktion auf eine notwendige Geschwindigkeitsanpassung vor einer Kurve. Demnach wurde die Fahrgeschwindigkeit bei ACC-Nutzung ca. fünf Sekunden später reduziert als dies beim manuellen Fahren der Fall war. Das Fahrverhalten lässt Rückschlüsse auf den Fahrerzustand in dem Umfang zu, als dass womöglich ein Müdigkeitsanstieg für das beobachtete Vigilanzdekrement verantwortlich war. Dieses Ergebnis steht im Einklang mit dem von Neubauer et al. (2014) berichteten Vigilanzeffekt, wonach signifikant längere Bremsreaktionen auf ein plötzliches Einscheren eines Fahrzeugs im Egofahrstreifen bei ACC verglichen mit der manuellen Kontrollbedingung ermittelt wurden. Im Mittel war die Bremsreaktion um ca. 0.5 Sekunden nach 42-minütiger ACC-Fahrt verzögert (Neubauer et al., 2014). Übereinstimmend vermerken Vollrath et al. (2011) und Neubauer et al. (2014), dass der ermittelte Workload bei den Überwachungstätigkeiten deutlich, teils signifikant niedriger war als beim manuellen Fahren. Damit stehen diese Beobachtungen in einer Reihe mit den Ergebnissen von De Winter et al. (2014). Ein Ausschnitt der umfangreichen Metaanalyse der Autoren über 32 Studien zeigte für manuelles Fahren und ACC-Nutzung (insgesamt 24 Studien), dass der mittlere, selbst beurteilte Workload von Fahrern in Level 0 bei 43.5 % und in Level 1 nur bei 38.6 % des maximal mittels NASA-TLX bestimmbaren Workloads lag (De Winter et al., 2014).

Bei Verwendung eines Level 2 wurden ähnliche, Vigilanzdekrement-typische Effekte von verschiedenen Autoren berichtet. Während einer Versuchsfahrt mit einem Tesla Model S erfolgte

eine vibrotaktile Stimulation am linken Arm der Probanden, auf welche sie schnellstmöglich mit einem Handtasterdruck reagieren sollten (Biondi et al., 2018). Interessanterweise wurden annähernd keine Reize verpasst, die Genauigkeit wird mit ca. 99.9 % angegeben. Allerdings wurde in der ersten Hälfte der etwa 50-minütigen Versuchsfahrt bei Verwendung der Level 2-Funktion noch um ca. 0.13 Sekunden schneller reagiert als in der zweiten Hälfte. In der manuellen Kontrollbedingung trat dieser Effekt nicht in Erscheinung. Damit war Biondi et al. (2018) eine Replikation der von Stapel et al. (2017) ebenfalls während einer Tesla Model S-Fahrt beschriebenen Ergebnisse grundsätzlich gelungen. Im Vergleich zu Biondi et al. (2018) bedienten sich Stapel et al. (2017) einem auditiven *Oddball-Paradigma*. Hierbei zeigten sich signifikant langsamere Reaktionen in der Level 2-Bedingung als in der manuellen Kontrollfahrt – jedoch nur bei normaler Verkehrsdichte und nur bei unerfahrenen Level 2-Nutzern. Sobald die Fahr-szenerie eine höhere Monotonie aufwies, verschwand dieser Effekt und es traten – unabhängig vom Automatisierungslevel – mehr Auslassungen auf. Diese erhöhten sich ebenfalls, wenn unerfahrene Level 2-Nutzer die Tesla-seitig angebotene Fahrfunktion verwendeten. Stapel et al. (2017) argumentieren, dass die Überwachungstätigkeit eine hohe Beanspruchung mit sich bringt, die durch die Komplexitätsreduktion innerhalb der Verkehrsumgebung auf ein geringeres Anforderungsniveau gesetzt wird. Mit anderen Worten schienen die Probanden mehr Ressourcen für die Vigilanzaufgabe erübrigen zu können, wenn die Verkehrssituation monoton ist. Die Tatsache, dass hiermit auch eine höhere Langeweile i.S. einer Unterbeanspruchung verbunden sein könnte (vgl. Abschnitt 2.2.3.1), wird leider von Stapel et al. (2017) nicht diskutiert. Bemerkenswert ist jedoch, dass die vorangegangene Erfahrung mit Level 2-Funktionen eine Auswirkung auf die Aufrechterhaltungsfähigkeit der Vigilanz haben kann. Hierzu passt, dass nur erfahrene Level 2-Nutzer einen signifikant geringeren Workload bei Level 2 verglichen mit Level 0 angaben. Leider haben Biondi et al. (2018) weder den Faktor „Monotonie“ noch „Expertise“ bei ihren Realfahrtversuchen berücksichtigt – insofern sind keine Vergleiche zu Stapel et al. (2017) diesbezüglich möglich.

Körber et al. (2015) haben sich ebenfalls eines auditiven Oddball-Paradigmas bedient, um Vigilanzeinbußen während einer 42.5-minütigen Level 2-Fahrt im Simulator sichtbar zu machen. Allerdings konnten sie keine signifikanten Änderungen von Reaktionszeiten ausmachen. Die Autoren erklären, dass womöglich eine zu geringe Aufgabenschwierigkeit hierfür verantwortlich war; dies untermauert die fast perfekte Antwortgenauigkeit der Probanden, da so gut wie keine Auslassungen oder falschen Alarmer zu verzeichnen waren. Weiter erklären Körber et al. (2015), dass eine gute (kompensatorische) Passung zwischen der vermeintlich einfachen Vigilanzaufgabe und der (noch) vorhandenen Leistungsfähigkeit der Probanden in der Fahrsitua-

tion bestand, sodass die Fahrer auf einem moderaten Arousalniveau gehalten wurden. Die geringen müdigkeitsbedingten Effekte bezogen auf die Lidschlussindikatoren legen diesen Schluss nahe. Damit wird an dieser Stelle ein grundlegendes methodisches Problem bei Ermüdungsmessung über die Zeit greifbar: Die Zuhilfenahme einer sekundären Vigilanzaufgabe im Experiment kann die primär durch die Automatisierung bestehende Vigilanzaufgabe konfundieren, sodass (erwartete) Ermüdungseffekte nur marginal in Erscheinung treten. Im schlechtesten Fall kann eine zusätzliche, sekundäre Vigilanzaufgabe bewirken, dass vollkommen unklar ist, welche Leistungseinbußen überhaupt gemessen werden. So bliebe die Frage offen, ob der Fahrerzustand, der für die Überwachungstätigkeit der automatisierten Fahrfunktion von Interesse ist, erfasst wird, oder der, der lediglich die Performanz in der hilfsweise integrierten (grundlagenpsychologisch orientierten) Vigilanzaufgabe abbildet.

In den Fahrversuchen von Stapel et al. (2017) war das auditive Oddball-Paradigma annähernd äquivalent zu Körber et al. (2015) parametrisiert – auch hier war es den Autoren nicht möglich, konsistente Ermüdungseffekte bei Level 2 in Relation zum manuellen Fahren in allen Bedingungen aufzuspüren. Diese Problematik kann beispielsweise mit den Untersuchungsergebnissen von Atchley, Chan und Gregersen (2014) erklärt werden, die unter Bearbeitung einer verbalen Nebenaufgabe beim manuellen Fahren einen Anstieg von Wachheit im EEG verzeichneten.

Die Tatsache, dass Biondi et al. (2018) eine andere Stimulusmodalität in der sekundären Vigilanzaufgabe verwendeten, ändert per se nichts am methodischen Defizit der Ermüdungsmessung. Zusätzlich „störten“ sie die Probanden während des Ermüdens, indem subjektive Abfragen zum Mind-wandering erfolgten. So ist es umso erstaunlicher, dass zumindest eine signifikant geringere Herzrate während Level 2 (verglichen mit Level 0) verzeichnet wurde. Hinsichtlich des Mind-wanderings konnte bei Level 2 kein Effekt festgestellt werden. Körber et al. (2015) zeigten hingegen eine signifikante Zunahme von Mind-wandering am Ende ihrer Level 2-Fahrt, verglichen mit dem Fahrtbeginn. Dies ist konsistent mit den von ihnen beschriebenen marginalen Ermüdungseffekten anhand des Lidschlussverhaltens mit zunehmender Fahrtdauer. Mit Verweis auf Schmidt et al. (2011), die mittels EEG eindrucklich zeigen konnten, welches Aktivierungspotential subjektive Abfragen während der Fahrt besitzen, ist die von Körber et al. (2015) verwendete Beschränkung der Abfrage auf Fahrtbeginn und -ende methodisch zu bevorzugen. Mitunter ist hierdurch auch das erfolgreiche Aufspüren von Mind-wandering nach Level 2-Nutzung zu erklären.

Schömig und Kaussner (2010) haben in einer Studienreihe ein Detektionsverfahren für Müdigkeitsfahrten validiert. In diesem Kontext untersuchten sie auch die Verwendung einer Level 2-Funktion mit wenigen Probanden in einer bis zu 150-minütigen Fahrt. Ziel war es, verschiedene Müdigkeitslevel anhand einer videobasierten Müdigkeitsdetektion (über Kopf- und Augenbewegungen) zu durchschreiten. Derweil störten sie die Testteilnehmer bewusst u.a. mit Übernahmeaufforderungen zum manuellen Fahren. Teilweise profitierten diese hiervon und ihre Müdigkeit reduzierte sich – dieser Effekt war jedoch insofern inkonsistent, als dass einige Probanden konstant müde blieben oder sogar noch stärker ermüdeten.

Besonders in Fahrsimulatorstudien wird dem Probanden häufig im Anschluss an die Müdigkeitsmessung ein kritisches Ereignis präsentiert, um zu untersuchen, ob und wie eine potenziell gefährliche Verkehrssituation von einem vermeintlich ermüdeten Fahrer gelöst werden kann. Analog zu Neubauer et al. (2014) scherte bei Saxby et al. (2013) ein Van vor dem Egofahrzeug der Probanden ein, nachdem diese 10 oder 30 Minuten in einem Level 2 oder manuell fuhren. Hervorzuheben sind hier die experimentelle Manipulation bzw. Instruktion in den Versuchsbedingungen: In Level 0 wurde den Versuchsteilnehmern eine fortwährende, intensive Regelungstätigkeit der Fahrzeugführung abverlangt. Permanente Einflüsse von Seitenwind sollten die Auswirkungen aktiver aufgabenbezogener Müdigkeit (vgl. Abschnitt 2.3.2) sichtbar machen. Hingegen wurden Probanden in Level 2 instruiert, die Automatisierung permanent auf Fehler hin zu überwachen (die aber nie in Erscheinung traten). Nachdem die Automatisierung nach 30 Minuten spontan abgeschaltet wurde (respektive kein Seitenwind mehr auftrat) entschieden sich die Autoren die Ermüdungsauswirkung anhand der SDLP über fünf 30-sekündige Zeitabschnitte zu beurteilen. Nach Anfangs höherer SDLP nach Level 2 stabilisierte sich die SDLP auf einem signifikant niedrigeren Niveau als nach Level 0 mit Seitenwind. Auf den ersten Blick könnte somit davon ausgegangen werden, dass die ermüdungsbedingten Einflüsse auf die Querregelung nach Level 2 geringer ausfallen als nach einer anstrengenden manuellen Fahrt. Betrachtet man jedoch die anschließenden Reaktionszeiten auf den Einscherer (eine *Time-to-Collision* ist leider nicht angegeben), so verzeichneten Saxby et al. (2013) signifikant verzögerte Brems- und Lenkreaktionen nach Level 2 verglichen mit Level 0 und Seitenwind. Dieser Effekt war nach 30 Minuten immens groß: Wo nach Induktion aktiver aufgabenbezogener Müdigkeit immerhin noch sechs Probanden eine Kollision vermieden – und damit auf vergleichbarem Niveau zur Kontrollbedingung ohne Seitenwind lagen – kollidierten alle Probanden nach der durch Level 2 hervorgerufenen passiven aufgabenbezogenen Müdigkeit mit dem Van. Die Autoren diskutieren, dass die SDLP womöglich kein Maß der Wahl zur Müdigkeitsdetektion in diesem Kontext darstellte. Einerseits sei hier angemerkt, dass insbesondere nach

manueller Fahrt mit Seitenwind erhebliche Interferenzen in der SDLP durch das vorherige Querregelungsverhalten vorhanden sein könnten. Andererseits verzeichneten Gonçalves et al. (2016) sehr wohl signifikant höhere Querschleunigungen müder Probanden nach vorheriger Level 2-Fahrt, die sogar im Vergleich zu Saxby et al. (2013) nur 17 Minuten dauerte.

Die Arbeitsgruppe um Kunderinger hat in einer Versuchsreihe Müdigkeitseffekte in Level 2 (verglichen mit Level 0) sowohl im Fahrsimulator als auch im Realfahrzeug untersucht (Kunderinger et al., 2018, 2020). In ihrer ersten Studie haben Kunderinger et al. (2018) eine Stichprobe von 30 Probanden zwei aufeinander folgende, monotone Fahrten absolvieren lassen. Dabei erlebten die Teilnehmer entweder zunächst eine 45-minütige manuelle Fahrt, gefolgt von einer 45-minütigen Level 2-Fahrt, oder umgekehrt. Während Level 2 sollten die Probanden die Automatisierung permanent überwachen und bereit sein, jederzeit die Fahrzeugsteuerung zu übernehmen. Alle fünf Minuten bewerteten die Teilnehmer ihre gegenwärtige Müdigkeit mittels KSS, die auf einer Tablet-Applikation in der Mittelkonsole dargestellt wurde. Besonderes Augenmerk richteten die Autoren darauf, dass die Aufforderung zur Müdigkeitsbewertung sehr implizit erfolgte: Die Skala erschien sehr langsam auf dem Tablet und wurde nach der Bewertung wieder ebenso langsam ausgeblendet – so sollte eine Interruption im Ermüdungsprozess bestmöglich vermieden werden. Neben dem Ermüdungsverlauf über die Zeit, der in Abbildung 13 dargestellt ist, fanden Kunderinger et al. (2018) einen großen Alterseffekt innerhalb ihrer Stichprobe. Junge Versuchsteilnehmer ermüdeten signifikant schneller als Ältere. Unabhängig vom Alter zeigte sich, dass Probanden während Level 2 signifikant schneller ermüden als beim manuellen Fahren.

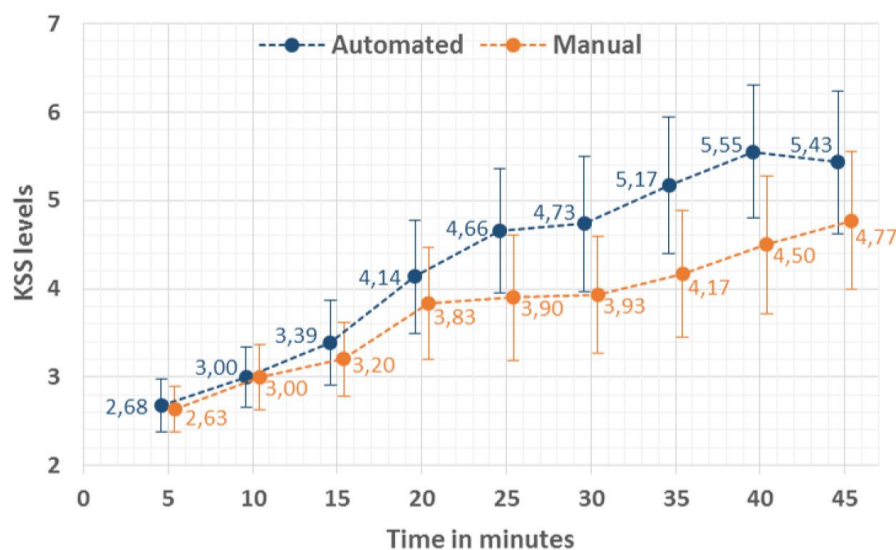


Abbildung 13: Müdigkeitsentwicklung anhand mittlerer KSS-Werte über die Zeit für manuelles und automatisiertes Fahren (Level 2). Die Fehlerbalken repräsentieren das 95 % Konfidenzintervall. Aus Kunderinger et al. (2018), S. 234.

In ihrer zweiten Studie verfolgten Kundinger et al. (2020) das Ziel, ihre im Fahrsimulator erzielten Ergebnisse in einem Realfahrzeug zu replizieren. Hierfür verwendeten sie ein mit Steuerungsrobotern ausgerüstetes Fahrzeug unter kontrollierten Bedingungen auf einer abgesperrten Teststrecke. Leider war aufgrund der Streckencharakteristik lediglich eine Fahrt mit 25 km/h möglich, was jedoch den Monotoniegrad verschärfte. Die Fahrdauer, die Instruktion sowie die Müdigkeitsbewertung lief nach selbem Schema wie in Kundinger et al. (2018) ab. Tatsächlich konnte der Effekt der Level 2-Automatisierung auf die Ermüdung repliziert werden (vgl. Abbildung 14). Im Vergleich zur ersten Studie zeigte sich kein monoton steigender, quasi-linearer Verlauf der Ermüdung über die Zeit. Stattdessen ermittelten Kundinger et al. (2020) ein maximales Müdigkeitslevel nach 35 Minuten, anschließend bleibt die Müdigkeit auf konstantem Niveau, bzw. fällt leicht ab. Die zusätzlich erhobene Herzrate erwies sich lediglich als sensitiver Müdigkeitsindikator für das jeweilige Automatisierungslevel, nicht aber für den Messzeitpunkt. Ferner fanden Kundinger et al. (2020) eine signifikante Korrelation zwischen Müdigkeitslevel und Vertrauen in die Automatisierungsfunktion nach der Level 2-Fahrt: Müdere Probanden besaßen ein höheres Funktionsvertrauen. Deskriptiv erreichten Probanden, die abends ab 17:30 Uhr getestet wurden, die geringsten KSS-Müdigkeitswerte. Diese Tendenz wird von den Autoren leider nicht weiter diskutiert, jedoch kann anhand dieser Beobachtung auch ein Einfluss schlafbezogener Müdigkeit, hier konkret zirkadiane und/oder chronotypische Faktoren, in der ganzheitlichen Müdigkeitsmessung vermutet werden.

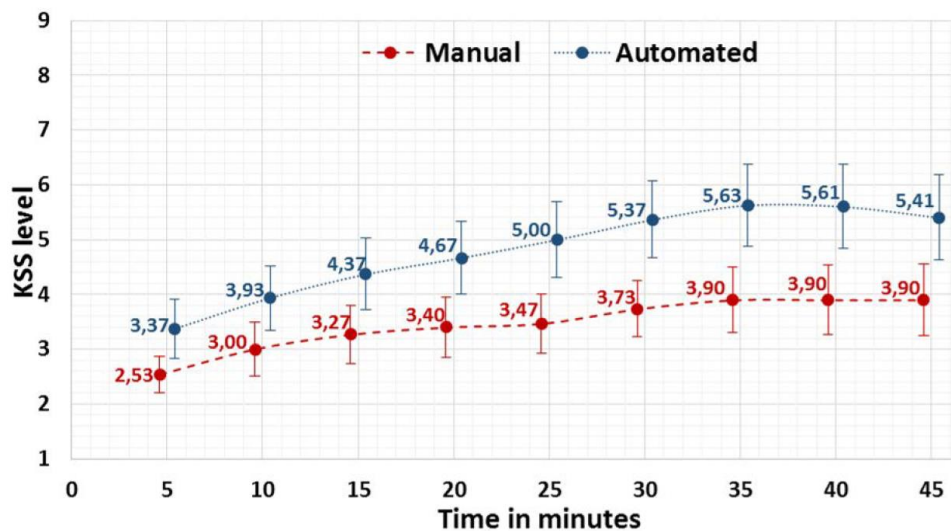


Abbildung 14: Müdigkeitsentwicklung anhand mittlerer KSS-Werte über die Zeit für manuelles und automatisiertes Fahren (Level 2). Die Fehlerbalken repräsentieren das 95 % Konfidenzintervall. Aus Kundinger et al. (2020), S. 375.

2.4.5.2 Studienlage für SAE Level 3

De Winter et al. (2014) haben nicht nur den mittleren, selbst beurteilten Workload von Fahrern in Level 0 (43.5 %) und in Level 1 (38.6 %), sondern auch für Level 3 in ihrer Metaanalyse bestimmt. Dieser lag demnach bei lediglich 22.7 % des maximal mittels NASA-TLX bestimmbaren Workloads. Zwar beschreiben De Winter et al. (2014) bei Level 3, dass sich Probanden häufiger mit fahrfremden Tätigkeiten (FFT) während automatisierter Phasen beschäftigen als in Level 1, was jedoch auch nicht verwundert, da diese Beschäftigung in Level 3 explizit möglich und auch so vorgesehen ist. Allerdings (vgl. Abschnitt 2.1) ist die Beschäftigung mit FFT gemäß SAE-Definition nicht verpflichtend. In ihrer Workload-Berechnung differenzieren De Winter et al. (2014) jedoch nicht zwischen Studien, in denen Level 3 mit oder ohne die Bearbeitung von FFT genutzt wurde. Die Übersicht anhand Tabelle 5 legt hingegen nahe, dass eine Unterscheidung entsprechend der Probandeninstruktionen in dieser Hinsicht absolut notwendig ist, da Müdigkeits- und vigilanztypische Effekte (dazu zählt auch der subjektive Workload) hiervon maßgeblich beeinflusst werden können.

Schömig et al. (2015) verglichen in einer experimentellen Studie die Müdigkeitsmuster während Level 3 ohne FFT und mit FFT (Quiz). Hierfür ließen sie 16 Probanden eine monotone, manuelle Fahrt im Simulator absolvieren. Über Lidschlussindikatoren war es möglich, vier vorab definierte Müdigkeitsniveaus zu unterscheiden. Sobald das jeweilige Niveau erreicht war, wurde für 15 Minuten mit Level 3 (oder zur Kontrolle weiter in Level 0) gefahren (hier Testphase genannt). Innerhalb des ersten 15-minütigen Blocks stieg die Fahrermüdigkeit in Level 3 ohne FFT ähnlich zu Level 0 an. Im weiteren Verlauf zeigte sich ein vergleichbares Muster analog zu den Level 2-Untersuchungen von Kundinger et al. (2020): Zwischen dem zweiten und dem vierten 15-minütigen Block zeichnete sich eine Art Müdigkeitssättigung ab – die Müdigkeit blieb auf hohem Niveau relativ konstant. Insgesamt wurde über alle Blöcke in Level 3 ohne FFT die größte Müdigkeitszunahme (verglichen mit Level 0 und Level 3 mit FFT) gemessen. Auch die Tatsache, dass nach einer Übernahmeaufforderung aus Level 3 (unabhängig vom Müdigkeitsniveau) ohne FFT ein deutlicher Aktivierungseffekt von Schömig et al. (2015) konstatiert wurde, unterstreicht die Ähnlichkeit zwischen Level 2 und Level 3 ohne FFT bezogen auf die Müdigkeitsentwicklung. Hingegen konnte über die gesamte Testphase (über alle vier 15-minütigen Blöcke) keine nennenswerte Müdigkeitsänderung während Level 3 *mit* FFT verzeichnet werden. Nach Übernahmeaufforderungen aus Level 3 mit FFT war es logischerweise auch nicht möglich, einen Aktivierungseffekt nachzuweisen, weil der hierfür benötigte vorherige Müdigkeitsanstieg erst gar nicht vorhanden war. Zur Veranschaulichung der Unterschiede zwischen Level 3 ohne und mit FFT sind in Abbildung 15 die von Schömig et al.

(2015) gemessenen Müdigkeitsentwicklungen während der Testphase und der Übernahme-phase dargestellt. Insgesamt werden die Ergebnisse der Studie dadurch getrübt, dass keine zu-fallskritische Analyse möglich war, da die Zellenbesetzung der einzelnen Bedingungen zu klein war. Folglich konnte lediglich deskriptiv analysiert werden. Ein größerer Stichprobenumfang hätte dieser Untersuchung einen hohen Mehrwert gebracht, zumal die Autoren mit hohen inter-individuellen Unterschieden im Ermüdungsverhalten gerechnet hatten.

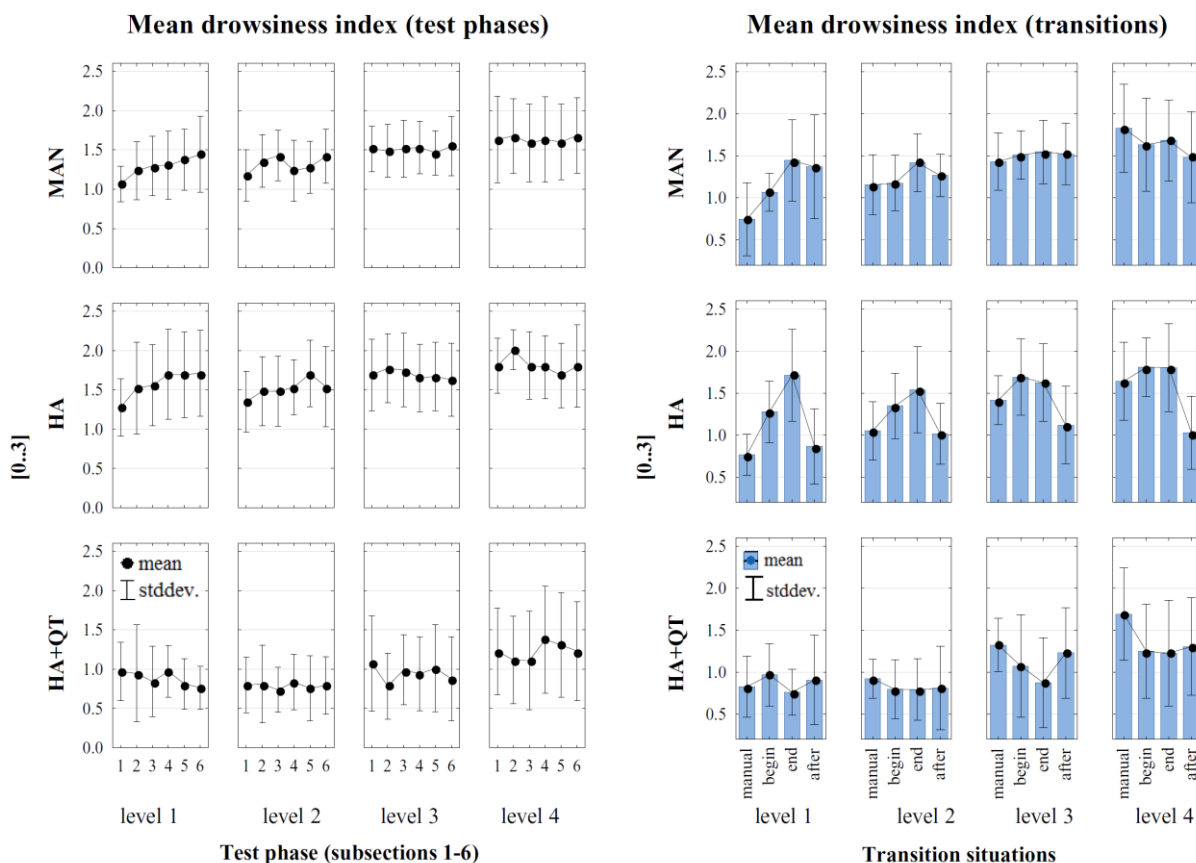


Abbildung 15: Links: Mittlere Müdigkeitsentwicklung über sechs 2.5-minütige Zeitbereiche innerhalb vier verschiedener Müdigkeitsniveaus (hier „level 1 bis 4“), ermittelt über Lidschlussindikatoren (hier als Indizes). Rechts: Mittlere Müdigkeitsentwicklung während der vorherigen Level 0-Fahrt (hier „manual“), zu Beginn der Testphase (hier „begin“), am Ende der Testphase (hier „end“) und nach Übernahmeaufforderung (hier „after“). Unterschieden wird Level 0 (hier „MAN“), Level 3 ohne FFT (hier „HA“) und Level 3 mit FFT (hier „HA+QT“). Aus Schömig et al. (2015), S. 6657.

Insofern lohnt ein Blick auf die experimentelle Untersuchung von Feldhütter et al. (2017), in der ebenfalls Level 3 ohne und mit FFT untersucht wurde, jedoch inferenzstatistische Analysen zulässig waren. Im statischen Fahrsimulator fuhren 15 Probanden Level 3 ohne FFT, weitere 15 Probanden mit FFT (SuRT). Sowohl nach 5 als auch nach 20 Minuten Fahrdauer mussten

die Teilnehmer die manuelle Fahrzeugsteuerung übernehmen. Neben einer signifikant schnelleren Übernahmezeit aus Level 3 ohne FFT heraus – was logisch erscheint aufgrund der eingesparten Abwendezeit von der FFT – fanden die Autoren eine signifikant schnellere Erstreaktion auf die Übernahmeaufforderung, wenn die Probanden lediglich 5 Minuten Level 3 fuhren verglichen mit 20 Minuten. Dieser Effekt ist zwar statistisch nachweisbar, jedoch von untergeordneter praktischer Relevanz, da der Reaktionsnachteil nach 20 Minuten im Mittel lediglich 90 ms betrug. Lohnender ist hingegen ein Blick auf das Kontrollblickverhalten der Versuchsteilnehmer während der 20-minütigen Level 3-Fahrt ohne FFT. Zu Beginn (nach 3 Minuten) fanden noch signifikant mehr Kontrollblicke auf die Fahrszenerie als nach 18 Minuten statt. Feldhütter et al. (2017) führen dies auf eine höheres Müdigkeitsniveau nach 18 Minuten zurück. Leider fehlen Informationen für die Level 3-Fahrt *mit* FFT (z.B. anhand der SuRT-Performanz nach 3 und 18 Minuten), um einen potenziellen Aktivierungseffekt der FFT fundierter nachweisen zu können.

Sowohl Vogelpohl et al. (2017) als auch Weinbeer et al. (2017) konzentrierten sich ausschließlich auf die Analyse des Fahrerzustands während Level 3 *ohne* FFT. Einige Ergebnisse lassen sich daher sehr gut vergleichen, unter der Berücksichtigung, dass Vogelpohl et al. (2017) einen statischen Fahrsimulator, hingegen Weinbeer et al. (2017) ein Wizard-of-Oz Fahrzeug im Realverkehr verwendeten. Hinsichtlich der zugrunde gelegten Müdigkeitsindikatoren gab es zwischen beiden Studien keine großen Unterschiede, da in beiden Fällen ein beobachterbasiertes Müdigkeitsrating zur Anwendung kam. Vogelpohl et al. (2017) berichten von einem deutlichen Müdigkeitsanstieg ihrer nicht schlafdeprivierten Versuchsteilnehmer während Level 3 ohne FFT bereits nach 20 Minuten Fahrdauer. Hier hatten ca. 20 % ein mittleres oder hohes Müdigkeitslevel erreicht, nach 45 Minuten stieg der Anteil auf etwa 50 % an. Verglichen hiermit konstatierten Weinbeer et al. (2017) auch ca. 20 % ihrer Probanden nach 20 Minuten ein mittleres, etwa 3 % ein hohes Müdigkeitslevel. Nach 45 Minuten war bei ca. 20 % ihrer Probanden ein mittleres Müdigkeitslevel, bei 40 % ein hohes Müdigkeitslevel erreicht. Die Ergebnisunterschiede zwischen beiden Studien sind letztlich marginal. Obwohl Vogelpohl et al. (2017) einen deutlich höheren Monotoniegrad in ihrer Untersuchung verwendeten als Weinbeer et al. (2017) (moderate Verkehrsdichte und Dunkelheit vs. normale Verkehrsdichte und Tageslicht) und die intrinsische Motivation höher sein sollte, im realen Straßenverkehr aus Sicherheitsgründen wacher bleiben zu müssen als im Fahrsimulator, sind solche Ableitungen hier nicht möglich. Vielmehr stützen sich die Ergebnisse beider Studien aus Validitätsperspektive gegenseitig. Allerdings können diese studienspezifischen Faktoren sehr wohl dafür verantwortlich sein, dass Vogelpohl et al. (2017) verschiedene Reaktionszeiteinbußen bei Übernahmeaufforderung nachweisen konnten (verglichen mit Reaktionszeiten nach manuellem Fahren) – Weinbeer et

al. (2017) dies jedoch verwehrt blieb. Besonders fällt auf, dass in beiden Experimenten zwischen 25 und 30 % der Teilnehmer nach mehr als 60 Minuten Fahrdauer noch nicht einmal ein mittleres Müdigkeitslevel erreichten. Hier bleibt die Frage offen, ob die individuellen Unterschiede im Fahrerzustand so groß oder die Indikatoren nicht sensitiv genug sind, die Ermüdung zu erfassen.

Im Gegensatz zu der zuvor vorgestellten Methodik haben sowohl Miller et al. (2015) als auch Naujoks et al. (2018) den Fahrerzustand während Level 3 *mit* FFT untersucht. Interessanterweise verglichen beide Studien die Müdigkeitsentwicklung mit einer Level 2-Nutzung, jeweils mit solidem Stichprobenumfang im Fahrsimulator. Große Kongruenz bestand ebenfalls beim verwendeten Müdigkeitsindikator, dem eine videobasierte Müdigkeitsbeurteilung zugrunde lag. Naujoks et al. (2018) orientierten sich dabei an der Kategorisierung des Lidschlussverhaltens von Wierwille und Ellsworth (1994), wohingegen Miller et al. (2015) sich auf die Identifikation eines hohen Müdigkeitsniveaus beschränkten (mindestens zweimal gähnen (Zeiteinheit unklar), und/oder Lidschluss > 5 s). Die größten experimentellen Unterschiede bestanden in der Fahrdauer: Miller et al. (2015) arbeiteten lediglich mit drei 8.5-minütigen automatisierten Phasen, Naujoks et al. (2018) ließen ihre Probanden entweder 50 oder sogar 110 Minuten am Stück automatisiert fahren. Zwischen beiden Fahrdauern konnten Naujoks et al. (2018) für keine abhängige Variable signifikante Unterschiede feststellen. Hingegen konnten sie in einem Regressionsmodell zeigen, dass sowohl die fremdbeurteilte Müdigkeit, der Workload als auch die Motivation der Versuchsteilnehmer signifikante Prädiktoren für die Reaktionszeit auf ein kritisches Ereignis (Einscherer vor dem Egofahrzeug) nach Level 2 darstellen, nicht jedoch nach Level 3 mit FFT (Smartphone und Zeitschriften). Dieses Ergebnis steht im Einklang mit der Forschung von Berghöfer et al. (2018) in einem Wizard-of-Oz-Fahrzeug im realen Straßenverkehr, die ebenfalls auf beobachterbasierte Müdigkeitsbewertung setzten: Die Müdigkeit der Probanden besaß in ihrem Regressionsmodell keine signifikante Prädiktion für die Übernahmezeit aus Level 3 mit diversen angebotenen FFT.

Die Probanden von Naujoks et al. (2018) übernahmen die Fahrzeugsteuerung nach Level 2 mit höherer Kritikalität – beispielsweise wurde im Anschluss an Level 2 eine niedrigere mittlere minimale TTC ermittelt als nach Level 3 mit FFT, also eine verlängerte Reaktionszeit. Naujoks et al. (2018) folgern, dass sich die Fahrer nach Level 3 deutlich entspannter auf die Übernahme vorbereiten konnten und dementsprechend souveräner auf den Einscherer reagierten.

Miller et al. (2015) berichteten keine nennenswerten Unterschiede nach Level 2 respektive Level 3 mit FFT (Lesen und Videosequenz) auf einen Einscherer vor dem Egofahrzeug (nach

dreimal 8.5 Minuten Fahrt). Allerdings haben die Autoren zusätzlich die Reaktionszeit der Fahrer auf die Übernahmeaufforderung ausgewertet und festgestellt, dass nach Level 2 signifikant schneller reagiert wurde als nach Level 3 mit FFT. Nach dieser verhältnismäßig kurzen Fahrdauer sollte jedoch nicht von einem müdigkeitsbedingten, sondern vornehmlich von einem ablenkungsbedingten Effekt verminderter Reaktionszeit ausgegangen werden. Hierfür spricht ebenfalls, dass Miller et al. (2015) sowohl nach der zweiten als auch nach der dritten 8.5-minütigen automatisierten Phase während Level 2 eine signifikant höhere Müdigkeit als während Level 3 mit FFT nachweisen konnten. Dieses Resultat untermauert die Ergebnisse von Schömig et al. (2015), wonach die Bearbeitung von FFT während Level 3 ein aktivierendes, müdigkeitsreduzierendes Potential besitzt. Weinbeer, Bill, Baur und Bengler (2018) konnten darüber hinaus zeigen, dass Fahrer eine hohe Akzeptanz besitzen, mit Hilfe der Bearbeitung von FFT die Ausbildung passiver aufgabenbezogener Müdigkeit zu unterdrücken.

Aus diesem Grund sollen an dieser Stelle die Studien von Jarosch et al. (2017, 2019a) betrachtet werden, die die Auswirkung der FFT-Charakteristik auf die Ermüdung während Level 3 gezielt untersuchten. Beide Studien besitzen die Gemeinsamkeit, zwei verschiedene FFT hinsichtlich ihres Ermüdungspotenzials zu vergleichen. Hierfür entwickelten Jarosch et al. (2017) die sog. „PqpD-Aufgabe“, um eine monotone Vigilanzaufgabe als FFT zur Verfügung zu haben. Die Buchstaben „P“, „q“, „p“ und „d“ wurden wiederholt in randomisierter Reihenfolge auf einem berührungsempfindlichen Display präsentiert, mit der Instruktion, das Display zu berühren, sobald das „p“ erschien. In Jarosch et al. (2017) wurde diese Aufgabe während der Level 3-Fahrt im Vergleich zu einer potenziell aktivierenden FFT (Quiz) im Fahrsimulator eingesetzt. Zu verschiedenen Messzeitpunkten wurde die subjektive Selbsteinschätzung der Müdigkeit mittels KSS und der objektive PERCLOS-Wert erhoben, siehe Abbildung 16. Es ist anhand beider Indikatoren deutlich erkennbar, dass die monotone FFT (signifikant) höhere Müdigkeit erzeugt, wobei das aktivierende Quiz die Fahrer fortwährend auf einem niedrigen Müdigkeitsniveau hält. Des Weiteren wird das Aktivierungspotenzial einer Übernahmeaufforderung zum manuellen Fahren (TOR) (vgl. Schömig et al., 2015) sichtbar: Nur bei Bearbeitung der monotonen FFT wird die hierdurch erzeugte Ermüdung kurzzeitig kompensiert.

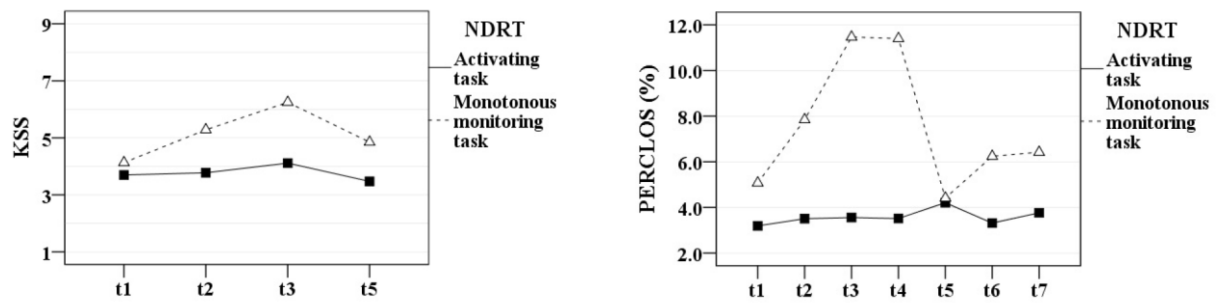


Abbildung 16: Links: Mittlere KSS-Werte abhängig von der Fahrdauer für die jeweilige FFT (hier NDRT). Die Messzeitpunkte sind wie folgt bestimmt: t1 = min 2-3, t2 = min 10-11, t3 = min 18-19, t4 = min 22-23, t5 (TOR) = min 24-25, t6 = min 25-26, t7 = min 26-28. Rechts: Mittlere PERCLOS-Werte abhängig von der Fahrdauer für die jeweilige FFT (hier NDRT). Aus Jarosch et al. (2017), S. 323.

In Jarosch et al. (2019a) wurden zwei grundlegende methodische Anpassungen durchgeführt. Einerseits fand die Untersuchung im realen Straßenverkehr in einem Wizard-of-Oz-Fahrzeug statt, andererseits wurde die vordefinierte aktivierende FFT durch eine frei wählbare FFT ersetzt. Die Fahrdauer wurde verlängert, eine Übernahmeaufforderung zum manuellen Fahren fand nicht mehr statt. Entsprechend Abbildung 17 wird deutlich, dass Jarosch et al. (2019a) die durch die verschiedenen FFT hervorgerufenen Müdigkeitssignaturen aus Jarosch et al. (2017) im Realverkehr replizieren konnten.

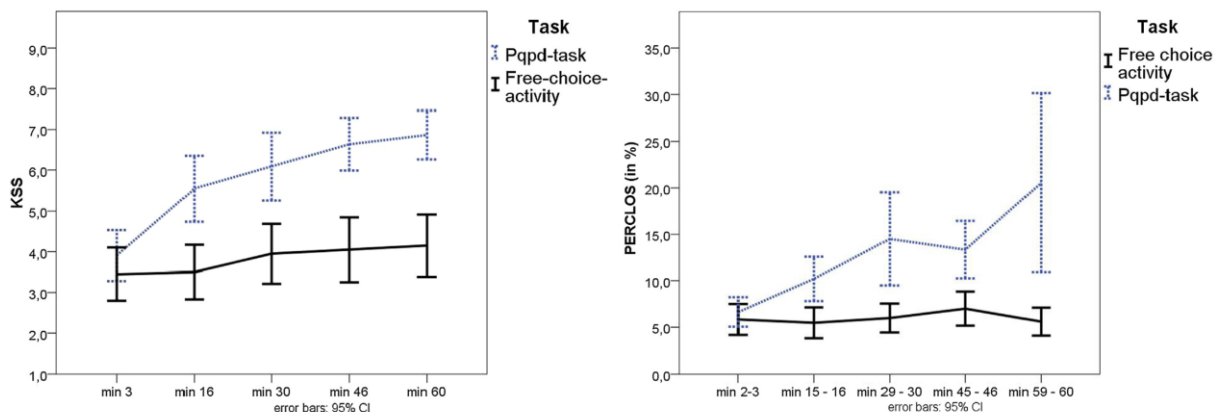


Abbildung 17: Links: Mittlere KSS-Werte abhängig von der Fahrdauer für die jeweilige FFT (hier Task). Aus Jarosch et al. (2019a), S. 302.⁶ Rechts: Mittlere PERCLOS-Werte abhängig von der Fahrdauer für die jeweilige FFT (hier Task). Aus Jarosch et al. (2019a), S. 301.⁶

⁶ Abgedruckt aus „Effects of Non-Driving Related Tasks in Prolonged Conditional Automated Driving – A Wizard of Oz on-Road Approach in Real Traffic Environment“ von Jarosch, O., Paradies, S., Feiner, D., & Bengler, K., erschienen in „Transportation Research Part F: Traffic Psychology and Behaviour, 65, 292-305“, Copyright 2019, mit Erlaubnis von Elsevier.

Insgesamt kann anhand der Studien von Jarosch et al. (2017, 2019a) geschlossen werden, dass nicht jede FFT das von Miller et al. (2015) oder Schömig et al. (2015) postulierte Aktivierungspotenzial besitzt, sondern die Charakteristik der Aufgabe (der Monotoniegrad) hierüber entscheidet. Sobald eine motivierende FFT bearbeitet wird, besteht Einigkeit der Autoren, damit die Ausbildung passiver aufgabenbezogener Müdigkeit während Level 3 zu unterdrücken.

2.4.6 Zusammenfassung Fahrerzustand und offene Forschungsfragen

Auf Basis der Literaturrecherche wurde ersichtlich, dass das Konstrukt des Fahrerzustands beim automatisierten Fahren verschiedene Facetten umfassen kann – Müdigkeit ist davon nur ein Teilgebiet, mit erheblichem Forschungsbedarf. Große Einigkeit besteht in der Forschungsgemeinschaft, dass das Ermüdungsverhalten höchsten intra- und interindividuellen Unterschieden unterliegt. Diverse Faktoren wie die Fahrdauer, die Zuverlässigkeit der Regelungstätigkeit, das Alter des Fahrers, die Fahrerfahrung mit der Arbeitsweise der Funktion, die Ausbildung von Kinetose und die Art der Beschäftigung mit FFT können ernstzunehmende Einflussgrößen für die Ermüdungsqualität darstellen. Für eine Studienplanung sollten auch zirkadiane Faktoren oder z.B. die Koffeineinnahme berücksichtigt werden. In einigen Untersuchungen war es beispielsweise nicht möglich, bei allen Fahrern einer Stichprobe konkrete Ermüdungsmuster festzustellen, da per se keine Müdigkeit messbar war. Obwohl in gegenwärtigen Forschungsarbeiten immer wieder darauf hingewiesen wird, dass diese intra- und interindividuellen Unterschiede im Ermüdungsverhalten bestehen und eine erhebliche Fehlervarianzquelle darstellen, wurde bisweilen die Quantifizierung der verschiedenen Einflussgrößen und ihr Zusammenwirken in der Forschung ausgespart. Dies gilt einerseits für Level 2, andererseits aber auch beim Vergleich zwischen Level 2 und Level 3.

Verschiedene Autoren haben unabhängig voneinander großes Bedenken, ob die kontinuierliche Überwachung einer automatisierten Fahrfunktion gemäß Level 2 vom Menschen sicher bewerkstelligt werden kann, ohne bereits nach kurzer Zeit zu ermüden. Sowohl Erkenntnisse aus grundlagenpsychologischen Experimenten als auch aus anwendungsorientierten Fahrversuchen geben deutliche Hinweise auf die Problematik des Vigilanzdekrements. Demnach können Fahrer in Level 2 nach etwa 15 Minuten Überwachungstätigkeit deutliche Anzeichen erhöhter Müdigkeit zeigen, die im weiteren Verlauf bis nach etwa 30 bis 40 Minuten fortwährend zunimmt. Anschließend kann eine Müdigkeitssättigung in Erscheinung treten. Ein ähnliches Muster gilt ebenfalls für Level 3 – allerdings nur, wenn keine FFT bearbeitet werden. Ohne FFT nähert sich das Fahrerverhalten in Level 3 nämlich dem in Level 2 an: Obwohl eine permanente

Überwachungstätigkeit in Level 3 nicht erforderlich ist, können Fahrer ohne FFT-Beschäftigung die Regelungstätigkeit der Funktion und die Verkehrsumgebung beobachten. Generell besitzen FFT nur dann ein aktivierendes Potential, wenn ihre Merkmale motivierend für den Fahrer sind. In diesem Fall können Ermüdungserscheinungen während Level 3 mittels FFT effektiv verhindert werden. Diese Ergebnisse gelten sowohl in simulierten Prüfungsumgebungen als auch auf der Teststrecke oder im Realverkehr.

Es fällt auf, dass bei der Überwachung von automatisierten Fahrfunktionen ein systematisch niedrigerer Workload der Fahrer vorzuherrschen scheint als beim manuellen Fahren. Dieses Resultat der Literaturanalyse lässt neben dem zuvor thematisierten anstrengenden Charakter von Vigilanzaufgaben auf eine vorwiegende Unterbeanspruchung der Fahrer in Level 2 und Level 3 ohne FFT schließen. Allerdings ist damit der potentielle Dualismus zwischen Unter- und Überbeanspruchung nicht abschließend geklärt: Aktuelle Studien fokussieren (wenn überhaupt) auf *einen* ausgewählten subjektiven Beanspruchungsindikator neben der primären Müdigkeitsmessung und diskutieren nicht das Zusammenwirken mehrerer Indikatoren. Beispielsweise wird entweder Stressempfinden operationalisiert *oder* der kognitive Workload erhoben. Sowohl für Level 2 als auch Level 3 zeigt sich insbesondere eine große Forschungslücke bei der Wechselwirkung zwischen Müdigkeit und Funktionsvertrauen: Obwohl eine gegenseitige Moderation der Konstrukte denkbar ist, wurde diese Forschungsfrage bis dato weder für Level 2 noch Level 3 aufgegriffen. Speziell in Level 2 könnte erhöhtes Vertrauen in die Steuerungstätigkeit mit Ausbildung von Langeweile einhergehen, und damit einen Redundanzgewinn für die Unterbeanspruchungshypothese liefern.

Grundsätzlich wäre für die Konzeption einer Level 2-Funktion von großem Interesse, ob Fahrer eine geringere Funktionszuverlässigkeit, d.h. mehr Automatisierungsfehler akzeptieren würden, oder hierdurch ihr Funktionsvertrauen auf ein zu niedriges Niveau sinken würde. Schließlich werden aus der Grundlagenforschung positive Aktivierungseffekte durch eine Erhöhung der Signalrate bei der Bearbeitung von Vigilanzaufgaben berichtet (vgl. Abschnitt 2.2.3). Mitunter könnte über eine geringere Zuverlässigkeit in Level 2 Müdigkeit reduziert und die Vigilanzleistung aufrechterhalten werden. Keine der zuvor analysierten Arbeiten hat diese Erkenntnisse aus der Grundlagenforschung bisher auf eine Level 2-Fahrt übertragen und die Hypothese entsprechend getestet.

Bei der Müdigkeitsmessung in Level 3 bedienen sich gegenwärtige Studien meist standardisierten oder nur einer einzigen ausgewählten FFT, statt die Bearbeitung natürlicher (vermeintlich motivierender) FFT auch im Wechsel zu gewähren. So besteht die Gefahr, dass lediglich

die durch die spezifische (ggf. monotone) FFT-Bearbeitung hervorgerufene Müdigkeit während Level 3 gemessen wird, und so das Entlastungspotential eines Level 3 (die Befreiung von der Überwachungsnotwendigkeit) in Relation zur Vigilanzaufgabe in Level 2 unsichtbar bleibt. Hier würde auch die Analyse des Kontrollblickverhaltens während Level 3 einen erheblichen Mehrwert bringen: Schließlich sollte der Fahrer in Level 3 davon profitieren, dass er keine Blicksequenzen zum Überwachen mehr benötigt, und so der vigilanzaufgabentypischen Ermüdung entgehen kann. Um dies zu gewährleisten, wäre jedoch ein umfassendes Funktionsvertrauen des Fahrers vonnöten. Dieser Forschungsgegenstand ist bis dato noch vollkommen unbeachtet.

Nach Beendigung einer Level 2- oder Level 3-Fahrt deutet die Literaturanalyse auf ein müdigkeitsabhängiges Übernahmeverhalten hin: Meist lassen die verwendeten Metriken auf eine Müdigkeitsreduktion nach erfolgter Übernahme schließen. Aktuell fehlt jedoch der unmittelbare Vergleich zwischen Level 2 und 3 bei äquivalenter Monotonie im Realfahrzeug und unter Zuhilfenahme desselben Müdigkeitsindikators. Folglich bleibt unklar, in welchem Ausmaß nach Level 2 eine Aktivierung im Fahrerzustand durch die Übernahme hervorgerufen wird: Es wird vermutet, dass nach Level 2 das Aktivierungspotential von Übernahmeaufforderungen höher ausfallen wird als nach Level 3, wenn bereits während der Level 3-Nutzung die Ausbildung von Müdigkeit unterdrückt wird. Erkenntnisse in dieser Hinsicht liegen jedoch noch nicht vor.

Obwohl von subjektiven Selbstbeurteilungen der Müdigkeit in der Literatur an verschiedenen Stellen abgeraten wird, wird aus Studien, die diese Methode dennoch anwendeten, deutlich, dass auch hiermit eine zuverlässige Messung der Müdigkeit im automatisierten Kontext möglich sein kann. Fremdbewertungen der Müdigkeit (mittels Videoanalyse) oder über Lidschlussalgorithmen können ebenfalls mit einer hohen Validität überzeugen. Generell fehlt jedoch ein abschließender Nachweis mittels objektiver Müdigkeitsdetektion, dass sich der Fahrerzustand während Level 2 in einer Realfahrt kongruent zum hinlänglich bekannten Vigilanzdekrement über die Zeit verhält. Insbesondere ist bis dato offen, ob das vermeintlich höchst valide Instrument des EEG für eine fundierte Müdigkeitsabschätzung beim automatisierten Fahren geeignet ist. Damit die dabei ermittelte Ermüdung auch tatsächlich auf die entsprechende experimentelle Manipulation zurückgeführt werden kann, bedarf es einer unmissverständlichen Instruktion der Probanden über die Funktionsgrenzen der Automatisierung und die ihnen damit zustehenden Aufgaben. Dies wurde in bestehenden Forschungsarbeiten häufig unzulänglich geleistet oder aufgrund zusätzlicher experimenteller Manipulationen entsprechend konterkariert, wodurch die Studienergebnisse verzerrt sein können.

Entsprechend der noch unbeantworteten Forschungsfragen wird im nachfolgenden Abschnitt 2.5 die Zielsetzung vorliegender Arbeit definiert.

2.5 Zielsetzung der Arbeit

Die vorliegende Arbeit beansprucht die zuvor identifizierten Forschungslücken zu schließen. Vorrangiges Ziel ist es, die Machbarkeit von objektiver Müdigkeitsmessung mittels EEG anhand einer realen Level 2-Fahrt zu prüfen und die Methodik unter kontrollierten Experimentalbedingungen zu validieren. Während der Bearbeitung dieser Vigilanzaufgabe soll die Müdigkeitsentwicklung von Fahrern kontinuierlich detektiert und unter Berücksichtigung verschiedener intra- und interindividueller Einflussgrößen analysiert werden. Es soll speziell geklärt werden, ob vornehmlich Unter- oder Überbeanspruchung für das Vigilanzdekrement während einer Level 2-Nutzung verantwortlich ist.

Die Arbeit möchte ferner Lösungen zur Aufrechterhaltung eines vorteilhaften Fahrerzustands beim automatisierten Fahren aufzeigen. Einerseits wird hierfür das Potential einer geringen Level 2-Funktionszuverlässigkeit erforscht. Andererseits soll ein Ansatz über die Entlastung von der Überwachungsnotwendigkeit gefunden werden, indem eine Level 2-Fahrt mit einer Level 3-Fahrt unter Bearbeitung motivierender FFT verglichen wird. Insbesondere interessieren die Unterschiede im Ermüdungsverhalten über die Fahrdauer sowie bei Übernahme der manuellen Fahrzeugsteuerung. Abschließend soll herausgefunden werden, welcher Zusammenhang zwischen der Müdigkeit des Fahrers und seinem Funktionsvertrauen besteht, um hierüber zusätzliche Stellgrößen zur Wahrung eines vorteilhaften Fahrerzustands zu identifizieren. In den nachfolgenden Abschnitten sind die Forschungsfragen formuliert, die auf Basis der vorangegangenen Literaturanalyse in zu testende Hypothesen für zwei Experimente überführt werden konnten.

2.5.1 Fragestellungen Experiment 1

1. Ist die Machbarkeit von Müdigkeitsmessung mittels EEG in einer automatisierten Realfahrt gegeben?
 - Die EEG-basierte Müdigkeitserfassung über Alpha-Spindeln sollte einen validen Müdigkeitsindikator darstellen und verglichen mit der videobasierten Müdigkeitsbeurteilung ein äquivalentes Ermüdungsverhalten abbilden. Nach dem Fahrversuch wird generell (subjektiv wie objektiv) von einer höheren Müdigkeit als vor dem Fahrversuch ausgegangen.

- Es wird erwartet, dass sich Fahrer entsprechend ihrer individuellen Ermüdung über Alpha-Spindeln klassifizieren lassen. Eine hohe Ermüdung sollte mit einer höheren Alpha-Spindelrate beziffert werden können als eine niedrige Ermüdung.
2. Wie entsteht und verläuft Müdigkeit beim Überwachen einer automatisierten Fahrfunktion (Level 2) über eine Dauer von ca. 60 Minuten?
- Es wird erwartet, dass Fahrer eine kontinuierliche Ermüdung über die gesamte Fahrdauer zeigen, wobei bereits nach 15 Minuten Level 2 eine ausgeprägte Müdigkeit vorliegen sollte.
3. Welche Merkmale von Beanspruchung, die typischerweise bei Vigilanzaufgaben in Erscheinung treten, treten ebenfalls bei Level 2-Nutzung auf?
- Durch die Level 2-Fahrt sollte sowohl eine Unter- als auch Überbeanspruchung hervorgerufen werden. Insofern kann Müdigkeit nicht nur durch geringen Workload i.S. eines zu niedrigen Arousal-Levels, sondern auch durch zu hohen Workload im Zusammenhang mit erhöhtem Stress auftreten. Dementsprechend müsste einerseits mit höherem Funktionsvertrauen aufgrund höherer Langeweile eine höhere Müdigkeit zu beobachten sein. Ferner legt die Studienlage den Schluss nahe, dass eine geringere kognitive Beanspruchung bei ermüdeten Probanden nach Level 2 nachweisbar sein sollte. Andererseits wird erwartet, dass nach der Level 2-Fahrt ein höheres Stresserleben als vor der Fahrt messbar ist.
4. Welche Auswirkung hat die erlebte Funktionszuverlässigkeit auf die Ermüdung und die subjektiven Bewertungen? Kann sich eine unzuverlässigere Funktion vorteilhaft in Hinblick auf den Fahrerzustand erweisen, weil Automatisierungsfehler aktivierend wirken?
- Eine hohe Funktionszuverlässigkeit – das heißt seltene Automatisierungsfehler – sollten eine höhere Ausbildung von Müdigkeit hervorrufen als eine niedrige Funktionszuverlässigkeit, da höheres Stresserleben verbunden mit einem erhöhten Workload (i.S. einer Überbeanspruchung) auftreten kann. Des Weiteren sollte eine hohe Funktionszuverlässigkeit nicht nur ein höheres Vertrauen, sondern ebenfalls eine höhere Ermüdung aufgrund höherer Langeweile (i.S. einer Unterbeanspruchung) bewirken.
 - Analog wird erwartet, dass bei geringerer Steuerungsgüte der Level 2-Funktion eine niedrigere Ermüdung zu beobachten sein müsste.
 - Reaktionen auf seltene Automatisierungsfehler sollten langsamer erfolgen als Reaktionen auf häufige Automatisierungsfehler. Die Reaktionszeit sollte mit höherer Müdigkeit länger ausfallen.

- Ob eine höhere Nutzungsbereitschaft und ein höherer Komfort bei einer höheren Funktionszuverlässigkeit in Erscheinung treten, kann a priori nicht beziffert werden, da – wie oben angenommen – womöglich auch die Müdigkeit mit höherer Funktionszuverlässigkeit ansteigt, was wiederum aversiv wahrgenommen werden kann.
5. In welcher Qualität und Quantität treten individuelle Unterschiede im Ermüdungsverhalten in Erscheinung?
- Es werden große individuelle Unterschiede erwartet. Mit verlängerter Schlafdauer vor der Fahrt oder vorheriger Koffeineinnahme sollte eine geringere Ermüdung zu beobachten sein. Bei jüngeren Fahrern wird von einer höheren Ermüdung als bei älteren Fahrern ausgegangen. Fahrer mit einer höheren Neigung zu Kinetose sollten mit höherer Wahrscheinlichkeit ermüden und dementsprechend ein negatives Befinden nach der Level 2-Fahrt äußern.

2.5.2 Fragestellungen Experiment 2

1. Lassen sich die Ergebnisse aus Experiment 1 in Bezug auf das Ermüdungsverhalten während Level 2 replizieren?
2. Welche Unterschiede im Ermüdungsverhalten und in den subjektiven Bewertungen existieren zwischen einer Level 2- und einer Level 3-Nutzung?
 - Es wird erwartet, dass eine Level 3-Fahrt Müdigkeitseffekten vorbeugen kann, d.h. die Ausbildung von Müdigkeit unterdrückt, wenn sich Fahrer mit motivierenden fahrfremden Tätigkeiten (FFT) beschäftigen, anstatt die Fahrt – wie in Level 2 – überwachen zu müssen.
 - In Level 3 sollte – zumal die Beschäftigung mit FFT möglich ist – ein höheres Vertrauen in die Fahrfunktion messbar sein als in Level 2, sowohl mit als auch ohne Fahrerfahrung.
 - Von Level 3 wird eine höhere Nutzungsbereitschaft erwartet, da sich a) die komfortbezogenen Vorteile von Fahrzeugautomatisierung erkennbarer abzeichnen (FFT-Beschäftigung ist möglich) und b) der Workload (i.S. einer Überbeanspruchung) verbunden mit dem Stresserleben in Level 3 niedriger ausfallen sollten als in Level 2.

3. Welche müdigkeitsbedingten Unterschiede existieren bei Übernahme der manuellen Fahrzeugsteuerung nach Level 2 verglichen mit Level 3? Welche Faktoren moderieren die Übernahmezeit?
 - Fahrer mit höherer Müdigkeit sollten die manuelle Fahrzeugsteuerung langsamer übernehmen als Fahrer mit niedrigerer Müdigkeit. Nach Level 2 sollten Probanden schneller die Übernahme bewältigen als nach Level 3, weil aus Level 3 zunächst die FFT-Beschäftigung beendet werden muss.
4. Welche Auswirkung hat eine Übernahmeaufforderung zum manuellen Fahren auf die Müdigkeit der Fahrer?
 - Es wird erwartet, dass durch den Aufgabenwechsel eine Durchbrechung der Monotonie stattfindet, die sich in einer Aktivierung des Fahrers nach der erfolgten Übernahme äußert.
 - Wenn das Ermüdungsverhalten während Level 3 erfolgreich unterdrückt wurde, sollte eine Übernahme nach Level 3 eine geringere Aktivierung als nach Level 2 bewirken.
5. In welchem Zusammenhang stehen Kontrollblicke, die während der FFT-Beschäftigung in Level 3 erfolgen, mit dem Funktionsvertrauen und dem Ermüdungsverhalten?
 - Eine höhere Anzahl von Kontrollblicken sollte ein niedrigeres Funktionsvertrauen anzeigen. Durch häufige Kontrollblicke, d.h. durch vermehrtes Überwachen der automatisierten Fahrt, wird eine höhere Ermüdung erwartet.

3 Experiment 1

3.1 Methode

Das erste Experiment zielte auf die Erfassung des Fahrerzustands, konkret der Müdigkeit, während einer automatisierten Realfahrt (Level 2) ab. Kernelement der Untersuchungsmethodik bildete der Einsatz eines Wizard-of-Oz-Forschungsfahrzeugs auf einer monotonen Teststrecke. Während einer Level 2-typischen Überwachungsaufgabe sollte mittels EEG über Alpha-Spindeldetektion menschliche Müdigkeit abgebildet werden, wobei besonderes Augenmerk auf die Manipulation der Level 2-Funktionszuverlässigkeit und der Messung individueller Unterschiede im Ermüdungsverhalten gelegt wurde. Ein videobasierter Ansatz zur Detektion von Müdigkeit wurde als zusätzliches Validitätskriterium herangezogen.

3.1.1 Versuchsplan

Das Ermüdungsverhalten sollte in Abhängigkeit von der automatisierten Fahrdauer (erste unabhängige Variable; UV 1) und der erlebten Funktionszuverlässigkeit (UV 2) untersucht werden. Probanden wurden in einem weiteren Schritt anhand ihrer individuellen physiologischen Ermüdung in „ermüdet“ und „nicht ermüdet“ geclustert (UV 3), um Auswirkungen auf weiteren abhängigen Variablen untersuchen zu können.

Die automatisierte Fahrdauer (UV 1) kann in der Theorie unendlich viele Faktorstufen annehmen. Um dieses zeitreihentypische Problem zu beherrschen, wurde die Abtastung der primären abhängigen Variablen – der EEG-Alpha-Spindelrate als objektives Müdigkeitsmaß – in minütigen Abständen festgelegt. Dieses Vorgehen findet sich beispielsweise bei Schmidt et al. (2011). Durch Aggregation der minütigen Alpha-Spindelraten auf acht konstruierte zeitliche Intervalle (im Folgenden Zeitbereiche genannt), wies die messwiederholte Variable „Fahrdauer“ dementsprechend acht Faktorstufen auf. Jedem Zeitbereich wurde eine Fahrstrecke von vier Runden auf dem Testfeld zugrunde gelegt – demnach umfasste jeder Zeitbereich eine Spanne von etwa acht Minuten oder geringfügig weniger: Eine Runde auf der Teststrecke (2.1 km) wurde bei einer Geschwindigkeit von 70 km/h (laut Tachometer), d.h. einer real angenommenen Geschwindigkeit von ca. 65 km/h, in ca. 116 Sekunden durchfahren. Dies entspricht fast zwei Minuten. Zu bedenken ist an dieser Stelle, dass die reale Rundenzeit aufgrund von Schwankungen in der Fahrgeschwindigkeit variierte. Für alle Analysen wird jedoch eine mittlere Rundenzeit über die gesamte Fahrdauer von zwei Minuten angenommen. Bei einer automatisierten Fahrt über 33 Runden, wobei die letzte Runde ausschließlich zum Verlassen

der Teststrecke diente, fuhren alle Probanden somit zwischen 60 und 64 Minuten ununterbrochen automatisiert.

Die zweite unabhängige Variable (UV 2), die „Funktionszuverlässigkeit“, wurde mit zwei Faktorstufen definiert: Die Probanden wurden einer von zwei Versuchsbedingungen (dem Erleben einer hohen oder einer niedrigen Funktionszuverlässigkeit) quasi-randomisiert zugewiesen (vgl. Abschnitt 3.1.2). In der Versuchsbedingung „niedrige Funktionszuverlässigkeit“ wurde in jeder vierten Runde (alle acht Minuten), d.h. in jedem Zeitbereich, ein einzelner Automatisierungsfehler präsentiert. Insgesamt sollten die Versuchsteilnehmer somit acht Fehler erleben. Die Fehlertypen wurden unterschieden in einer Verringerung der Fahrgeschwindigkeit unter 60 km/h (Typ a), einer Erhöhung der Fahrgeschwindigkeit über 80 km/h (Typ b), ein Verlassen des Fahrstreifens nach links (Typ c) und ein Verlassen des Fahrstreifens nach rechts (Typ d). Jeder Versuchsteilnehmer erlebte eine andere, zufällig permutierte Fehlerreihenfolge. Permutationsrestriktionen bestanden darin, dass jeder Fehlertyp genau zweimal auftreten sollte, jedoch nie derselbe Fehler unmittelbar nacheinander. Außerdem wurde festgelegt, innerhalb der ersten vier und innerhalb der letzten vier Fehlerpositionen alle Fehlertypen (a bis d) einmal erlebbar zu machen. Schließlich wurde an der siebten Fehlerposition das Auftreten des Fehlertyps (c) oder (d) definiert. Die Hälfte der Probanden erlebten hier somit (zufällig zugewiesen) eine Querablage nach links, die andere Hälfte nach rechts. Diese Restriktion sollte eine Vergleichbarkeit der Reaktionen auf den Automatisierungsfehler mit der Versuchsbedingung „hohe Funktionszuverlässigkeit“ erlauben. In dieser Bedingung wurde den Versuchsteilnehmern auch – und ausschließlich – an Position sieben ein Automatisierungsfehler Typ (c) oder (d) präsentiert, kongruent zur Hälfte Typ (c) und zur Hälfte Typ (d). Die Verankerung der UV 2 „Funktionszuverlässigkeit“ im Versuchsdesign ist in Abbildung 18 verdeutlicht. Abbildung 18 zeigt ebenfalls, an welchen Positionen die Fehlertypen auftreten konnten. Der erste Automatisierungsfehler in der niedrigen Zuverlässigkeitsbedingung fand an Position 1 (gerader Streckenabschnitt), der zweite Automatisierungsfehler an Position 2 (Kurve) statt, usw. Folglich fand der siebte Automatisierungsfehler wieder an Position 3 (gerader Streckenabschnitt) statt, analog auch in der hohen Zuverlässigkeitsbedingung.

Auf Basis der individuellen Alpha-Spindelrate sollten die Versuchsteilnehmer post hoc klassifiziert werden, ob generell von einer Ermüdung während des Versuchs auszugehen war, oder nicht. Die dritte unabhängige Variable (UV 3), die „Ermüdung“, umfasste demnach zwei Faktorstufen. Die Vorgehensweise dieser Ermüdungsklassifikation ist in Abschnitt 3.1.4.3 erläutert. Dem ersten Experiment lag demnach ein $8 \times 2 \times 2$ faktorielles Design mit Messwiederholung auf dem ersten Faktor zugrunde.

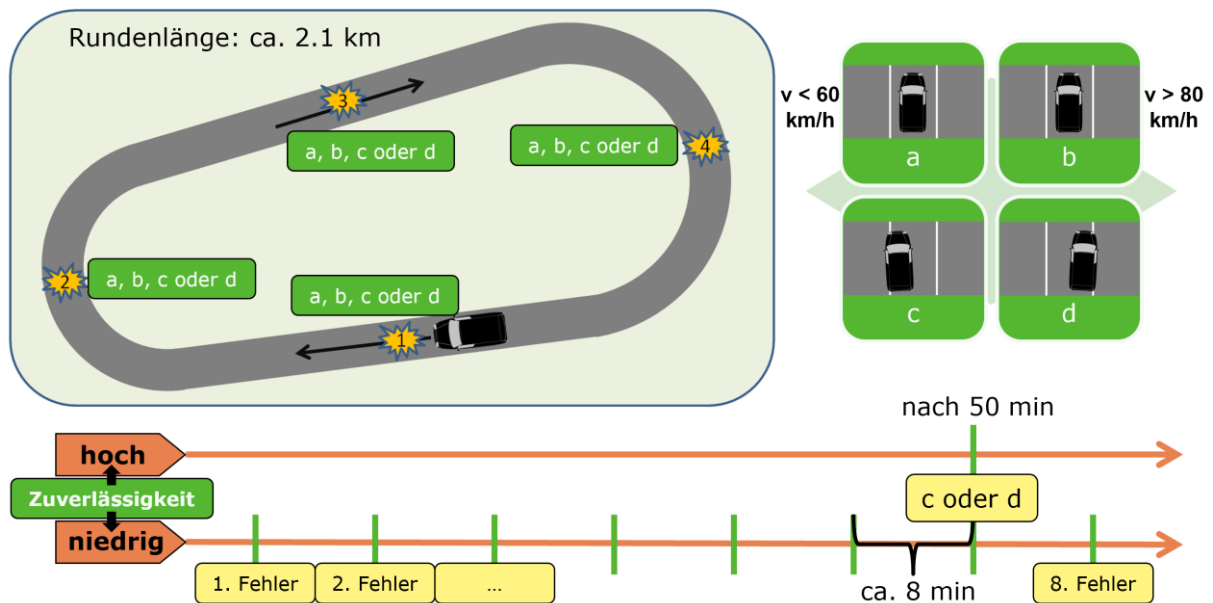


Abbildung 18: Versuchsdesign in Experiment 1. Oben links: Schematische Teststrecke, mit Fahrtrichtung des Wizard-of-Oz-Fahrzeugs und Positionen, an denen Automatisierungsfehler präsentiert wurden. Oben rechts: Fehlertypen (a-d). Unten: Vergleichende Darstellung der hohen und der niedrigen Funktionszuverlässigkeitsbedingung im zeitlichen Verlauf.

Die Auswirkungen der Manipulationen wurden mit Hilfe verschiedener abhängiger Variablen gemessen. Hierbei sind Subjektivurteile der Probanden und objektive Messmethoden zu unterscheiden. Die subjektiven Abfragen mittels Fragebogen (vgl. hierzu im Detail Abschnitt 3.1.4.4) wurden sowohl vor als auch nach der Versuchsfahrt durchgeführt, siehe Tabelle 6. Die objektiven abhängigen Variablen umfassten die Müdigkeitsdetektion mittels EEG, die Reaktionszeit auf Automatisierungsfehler sowie die videobasierte Müdigkeitsbeurteilung.

Tabelle 6: Subjektive abhängige Variablen in Experiment 1 inklusive Messzeitpunkt.

subjektive abhängige Variable	vor dem Fahrversuch	nach dem Fahrversuch
Müdigkeitseinschätzung	×	×
Kinetose	×	×
Vertrauen in die Automatisierung	×	×
Stressempfinden	×	×
Nutzungsbereitschaft / Komfort		×
Beanspruchung (Workload)		×
Schlafdauer vor der Versuchsfahrt		×
Koffeinkonsum am Morgen		×

3.1.2 Stichprobe

Insgesamt nahmen 38 Probanden an dem Versuch teil. Da bei zwei Probanden verwechselt, bzw. bei einem Probanden gar keine EEG-Daten aufgezeichnet wurden, wurden diese vollständig aus der Auswertung ausgeschlossen. Die Nettostrichprobengröße betrug daher $N = 35$. Hier von erlebten $n = 18$ eine hohe Funktionszuverlässigkeit und $n = 17$ eine niedrige Funktionszuverlässigkeit. Die Zuweisung zur jeweiligen Bedingung erfolgte so, dass eine annähernde Vergleichbarkeit zwischen den Bedingungen hinsichtlich Alter und Geschlecht der Probanden gewährleistet war. Der hohen Zuverlässigkeit wurden somit acht weibliche und zehn männliche Probanden im Alter zwischen 19 und 62 Jahren ($M = 36$, $SE = 3$) zugewiesen. Sie hatten im Mittel seit 19 Jahren ($SE = 3$) eine Pkw-Fahrerlaubnis und fuhren durchschnittlich ca. 12100 km ($SE = 2400$) im Jahr. In der niedrigen Zuverlässigkeitsbedingung wurden acht weibliche und neun männliche Probanden im Alter zwischen 19 und 60 Jahren ($M = 35$, $SE = 3$) untersucht. Sie hatten im Mittel seit 17 Jahren ($SE = 3$) eine Pkw-Fahrerlaubnis und fuhren durchschnittlich ca. 15800 km ($SE = 3700$) im Jahr.

3.1.3 Probandenakquise

Alle Probanden wurden per Zeitungsannonce auf die Studie aufmerksam gemacht. Das Inserat wurde in der Samstagsausgabe der Aachener Zeitung („Dürener“ und „Jülicher“ Ausgabe) mit folgendem Text geschaltet: „Die Bundesanstalt für Straßenwesen (BASt) sucht Personen, die Interesse haben, einmalig (Termine Mo. bis Fr.) an einer EEG-Studie zu automatisiertem Fahren teilzunehmen. Ort: Aldenhoven Testing Center, Zeitaufwand: ca. 4 Std., Vergütung: 80 EUR. Bei Interesse melden Sie sich bitte unter [...]“. Hieraufhin meldeten sich 105 Personen, wobei die meisten Interessenten zwischen 60 und 75 Jahren alt waren. Die Vertragsbedingungen der Unfallversicherung sahen jedoch eine Altersbegrenzung von 65 Jahren vor. Weitere Teilnahmevoraussetzungen an der Studie waren neben einer gültigen Fahrerlaubnis eine Mindestfahrleistung von 1000 km pro Jahr sowie keine Medikamenteneinnahme, die die Fahrtüchtigkeit beeinflussen kann. Um weitere Probanden (auch im Rahmen einer Nacherhebung) zu akquirieren, wurde die Anzeige drei Monate später erneut geschaltet, mit der inhaltlichen Anpassung, dass „noch Personen zwischen 18 und 65 Jahren“ für die Studie gesucht werden. Im Telefonat mit den Interessenten wurde erklärt, dass ihnen ein EEG zur Ableitung von Hirnströmen angelegt werden wird. Als Cover-Story wurde den Probanden erläutert, dass sie teilautomatisiertes Fahren in einem Forschungsfahrzeug der BASt erleben würden, um Rückmeldung über die Funktionsweise des Systems zu geben, damit eine Optimierung der Automatisierungsfunktion stattfinden kann. Ferner wurden organisatorische Aspekte besprochen, jedoch keine Informationen über den tatsächlichen Hintergrund der Studie preisgegeben. Falls Probanden vorab Fragen hatten, wurden Sie um Verständnis geben, dass aus methodischen Gründen erst nach der Versuchsteilnahme hierüber detaillierte Auskunft gegeben werden könne.

3.1.4 Versuchsmaterial

3.1.4.1 Teststrecke

Auf Basis verschiedener experimenteller Anforderungen an ein geeignetes Versuchsgelände wurde das Aldenhoven Testing Center (ATC) bei Aachen ausgewählt. Das Gelände verfügt u.a. über eine Teststrecke in Ovalform von ca. 2.1 km Länge, die drei Fahrstreifen besitzt. Zwei gerade Teilstücke sind über zwei Steilkurven mit unterschiedlichem Radius verbunden (vgl. Abbildung 19).

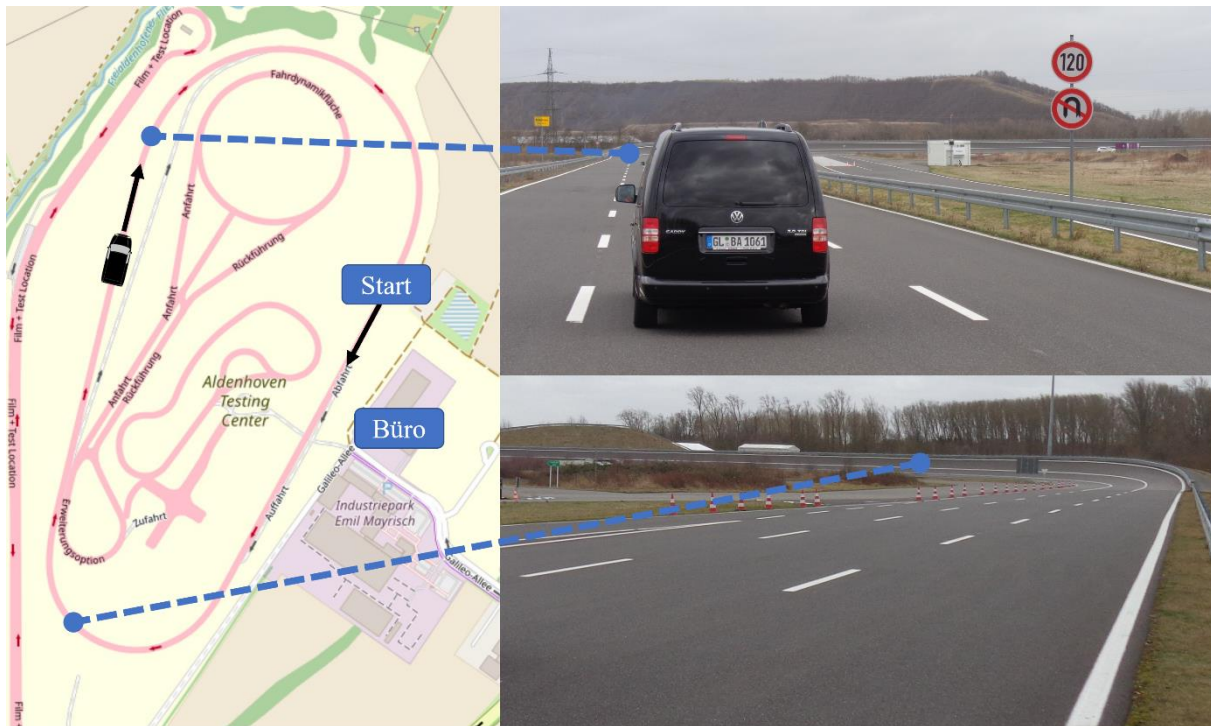


Abbildung 19: Kartendarstellung (Übersicht; Quelle: OpenStreetMap Deutschland) und ausgewählte, lokalisierte Ansichten der Teststrecke. Die Versuchsfahrt erfolgte in Pfeilrichtung.

Im mittleren Fahrstreifen mit einer Breite von 3.75 m konnte bis zu einer Fahrgeschwindigkeit von ca. 80 km/h eine annähernd fliehkraftfreie Fahrt für die Probanden gewährleistet werden. Alle Versuchsfahrten wurden unter exklusiver Nutzung durchgeführt, d.h. ausschließlich das hier verwendete Versuchsfahrzeug (siehe Abschnitt 3.1.4.2) befand sich während des Versuchs auf der Ovalbahn. Auch die Umgebung, die der Proband wahrnahm, zeichnete sich durch eine texturarme Landschaft aus (Bürogebäude, Bäume, einige Freileitungsmaste und eine versiegelte Mülldeponie). Da der Asphalt der Teststrecke zum Versuchszeitpunkt erst ca. zwei Jahre alt war, konnte auch bezogen auf Fahrbahnnunebenheiten ein hoher Monotoniegrad erreicht werden.

3.1.4.2 Versuchsfahrzeug

Der Fahrversuch wurde mit einem Forschungsfahrzeug der BASt durchgeführt. Dabei handelte es sich um einen „Volkswagen Caddy Maxi“, 2.0 TDI mit 140 PS und Automatikgetriebe aus dem Baujahr 2013 (vgl. Abbildung 20).



Abbildung 20: „Wizard-of-Oz-Fahrzeug“ der BAST.

Das Fahrzeug wurde vorab zu einem sog. „Wizard-of-Oz-Fahrzeug“ umgebaut. Nähere Erläuterungen zu diesem Fahrzeugkonzept finden sich in Abschnitt 2.4.4. Die Besonderheit des hier vorgenommenen Umbaus umfasste die vollständige Verdeckung des zweiten Fahrerplatzes gegenüber dem Probanden. Hierfür wurde der zweite Fahrerplatz als eigenständige Fahrerkabine im Fond vollständig von dem vorderen Probandenplatz abgetrennt (vgl. Abbildung 21).



Abbildung 21: Versuchsleiter („Wizard-Fahrer“) und Proband im „Wizard-of-Oz-Fahrzeug“.

Die Abtrennung erfolgte mittels einer stark getönten Glasscheibe. In Verbindung mit einer nahezu vollständigen Verdunkelung des Fonds war es dem Probanden nicht möglich, die Tätigkeiten des Wizard-Fahrers wahrzunehmen. Umgekehrt hatte der Wizard-Fahrer (bzw. Versuchsleiter) gute Sicht durch die Trennscheibe und die Windschutzscheibe hindurch auf die Fahrbahn. Im Fond des Versuchsfahrzeugs stand dem Versuchsleiter eine Einrichtung zur vollständigen Steuerung des Fahrzeugs zur Verfügung, siehe Abbildung 22.

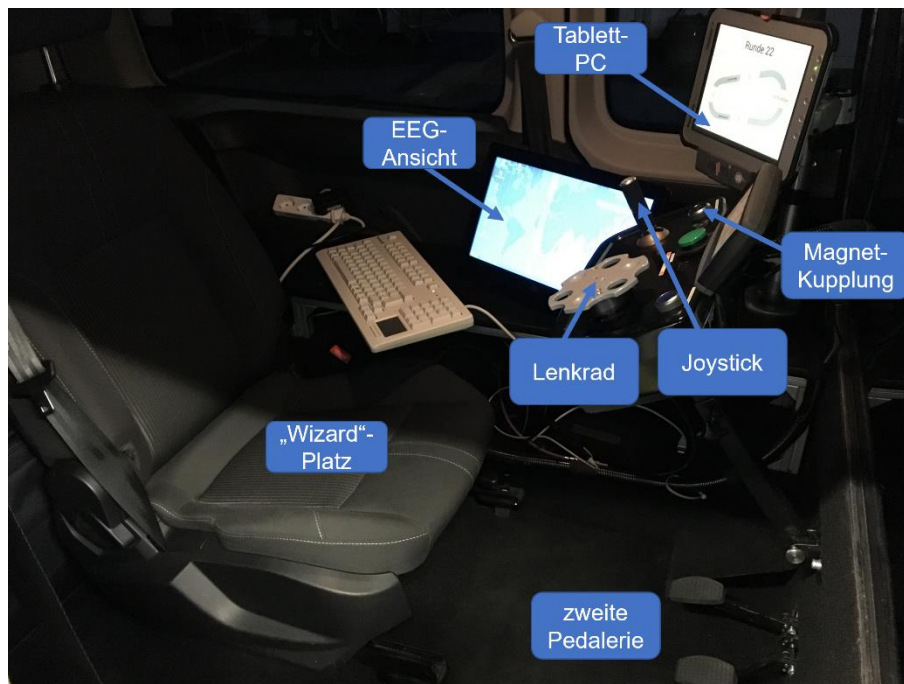


Abbildung 22: Beschriftete Übersicht des Versuchsleiterplatzes im Fond des „Wizard-of-Oz-Fahrzeugs“.

Die Fahrzeugsteuerung wurde durch ein nach ISO 26262 (2018) ASIL D abgesichertes „Drive-by-Wire-System“ der „Paravan GmbH“ für mobilitätseingeschränkte Personen umgesetzt. Auf diese Weise sind über Ausnahmegenehmigungen auch Forschungseinsätze des Fahrzeugs im öffentlichen Straßenverkehr möglich. Für die nachfolgenden Erläuterungen empfiehlt es sich, die Abbildung 22 zur Hand zu nehmen. Die Längsführung, d.h. die Gaspedal- und Bremspedalstellung konnte der Versuchsleiter entweder über eine Pedalerie (vermittelt über Seilzüge zur vorderen Pedalerie) oder über einen Joystick an einer zentralen Bedienkonsole variieren. Ebenfalls war es ihm durch ein kleines Lenkrad (aus seiner Perspektive rechts auf der Konsole) möglich, die Querführung, d.h. die Lenkung des Fahrzeugs zu kontrollieren. Hierfür musste vorab eine Magnetkupplung, die zusätzlich am Lenkgestänge verbaut wurde, geschlossen werden (in Abbildung 22 markierter Taster oberhalb des Joysticks auf der Bedienkonsole). Oberhalb der Bedienkonsole wurde über eine Folientastatur die Zündung betätigt und der Fahrzeugmotor gestartet. Die aktuelle Fahrgeschwindigkeit konnte durch die Trennscheibe hindurch direkt über den werkseitigen Tachometer des Fahrzeugs abgelesen werden. Links neben dem Versuchsleiterplatz wurde ein Tablet-PC montiert. Dieser zeigte die Übersicht auf die Teststrecke (vgl. Abschnitt 3.1.4.1), in der die zu fahrenden Automatisierungsfehler in der jeweiligen Runde eingebettet waren. So war es dem Wizard-Fahrer permanent einsehbar, in welcher Runde, in welchem Streckenbereich, er welchen Automatisierungsfehler zu fahren hatte. Nach jeder absolvierten Runde blätterte der Versuchsleiter eine Ansicht weiter. Auf dem Monitor links vom Fahrerplatz war die Ansicht des EEG-Messrechners, der sich auf dem Beifahrerplatz

neben dem Probanden befand, gespiegelt. Die Bedienung des EEG-Messrechners erfolgte über eine zusätzliche Tastatur links neben dem Versuchsleiterplatz. Auf diese Weise konnte der Versuchsleiter ohne Interaktion im vorderen Probandenbereich Messungen starten oder stoppen, die Plausibilität der EEG-Spannungsverläufe überprüfen oder auch die Videobilder der Probandenbeobachtungskameras ansehen.

Der vordere Fahrerplatz bot grundsätzlich alle werkseitig im Fahrzeug verbauten Funktionalitäten. Auf dem Beifahrersitz wurde zusätzlich das EEG-Aufzeichnungssystem festgeschnallt. Dieser Aufbau beinhaltete die Spannungsversorgung der aktiven EEG-Elektroden (eine genauere Beschreibung des verwendeten EEG findet sich in Abschnitt 3.1.4.3), die Spannungsversorgung des EEG-Messlaptops, den EEG-Messlaptop selbst und einen Videowandler (vgl. Abbildung 23).



Abbildung 23: Vorderer Probandenplatz mit EEG-Messrechner und Videokamera auf dem Dashboard.

Auf dem EEG-Messrechner wurden alle aufgezeichneten Daten zeitlich synchron erfasst und gespeichert. Bei der zuvor beschriebenen Messtechnik handelte es sich um Equipment, das der BAST im Rahmen des Verbundprojekts „Ko-HAF“ (Kooperatives, hochautomatisiertes Fahren; gefördert vom Bundesministerium für Wirtschaft und Energie (BMWi)) von der Daimler AG zur Verfügung gestellt wurde. Diese Messtechnik wurde analog in den Studien von Schmidt (2010) verwendet.

Um den Probanden herum wurden insgesamt vier Videokameras mit einer Auflösung von 320 x 240 Bildpunkten (VGA) platziert. Zwei Kameras, die auf dem Dashboard befestigt wurden, beobachteten den Probanden frontal (links bzw. rechts seitlich am Lenkrad vorbei, vgl. Abbildung 23 und 24). Mit Hilfe dieser Videobilder konnte die videobasierte Müdigkeitsbeurteilung (vgl. Abschnitt 3.1.7) erfolgen. Eine weitere Kamera (ebenfalls auf dem Dashboard fixiert) beobachtete die Fahrsituation frontal durch die Windschutzscheibe hindurch. Diese Videobilder dienten der nachträglichen Plausibilitätsprüfung der korrekten Zuordnung des real durchfahrenen Streckenabschnitts, der GPS-Position und des gefahrenen Automatisierungsfehlers. Ferner wurde eine vierte Kamera neben dem EEG-Messlaptop positioniert und auf Oberschenkel-Höhe des Probanden ausgerichtet (vgl. Abbildung 23). So konnte im Nachhinein ebenfalls überprüft werden, ob der Proband – wie es instruktionsseitig vorgesehen war – seine Hände während der automatisierten Fahrt auf den Oberschenkeln ablegte. Alle Kamerabilder wurden über den Videowandler an den EEG-Messlaptop übertragen.

Auf der Innenseite der Windschutzscheibe wurde dicht neben dem Innenspiegelfuß eine Sensorkamera der Firma „Mobileye“ befestigt, siehe Abbildung 24. Hierbei handelte es sich um eine „Mobileye 5“ mit einem optischen Sensor von 640 x 480 Bildpunkten, der bei einem horizontalen Öffnungswinkel von 38° die gesamte Fahrstreifenbreite erfassen konnte. So war die Kamera in der Lage, sowohl die Spurposition des Versuchsfahrzeugs innerhalb des eigenen Fahrstreifens als auch das Überschreiten der Fahrstreifenmarkierungen zu detektieren. Sobald eine Überschreitung nach links oder rechts erkannt wurde, wurde die Information über CAN-Bus an den EEG-Messrechner geschickt. Das Auslesen der Informationen auf dem CAN-Bus ermöglichte ein „Open-Data-Protocol“ der Firma „Mobileye“ für Forschungsvorhaben. Da die Sensorkamera ebenfalls die gegenwärtige Fahrgeschwindigkeit über den Fahrzeug-CAN-Bus erhielt, konnte indirekt über die Sensorkamera die aktuelle Fahrgeschwindigkeit über CAN-Bus an den EEG-Messrechner gesendet werden.

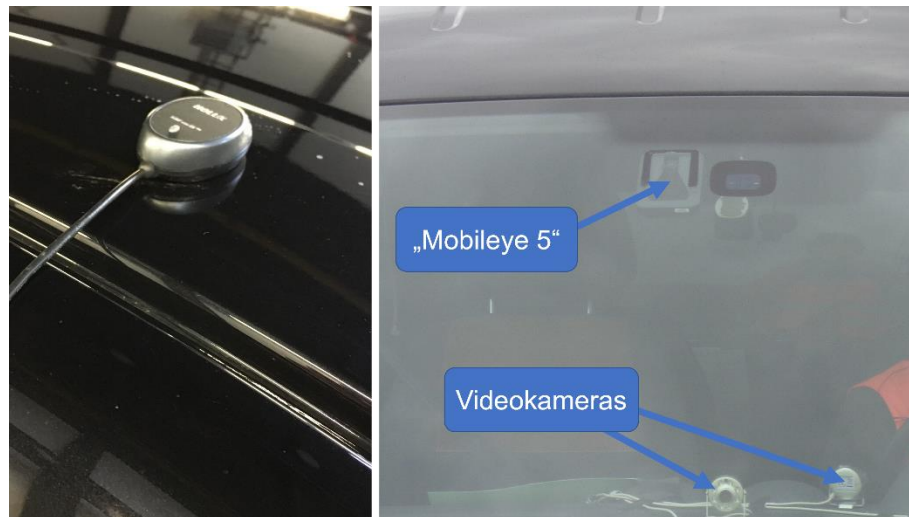


Abbildung 24: Links: GPS-Empfangsmodul; rechts: Sensorkamera „Mobileye 5“ und Videokameras auf dem Dashboard.

Zusätzlich wurde ein Handtaster am EEG-Messrechner angeschlossen. Dieser ermöglichte den Probanden das Erkennen von Ereignissen zu signalisieren. Der Taster besaß einen kreisförmigen Druckschalter (ca. 5 mm Durchmesser) und konnte an einem beliebigen Finger des Probanden mittels Kreppband befestigt werden. Außerdem fand ein GPS-Empfangsmodul (siehe Abbildung 24) Verwendung, das auf dem Fahrzeugdach über dem Beifahrerplatz magnetisch befestigt wurde und ebenfalls an den EEG-Messlaptop angeschlossen wurde.

3.1.4.3 Elektroenzephalographie (EEG)

Für die Ableitung des EEG wurde sich für die Verwendung aktiver Elektroden mit Impedanzwandlung entschieden. Das „actiCAP“-System (BrainAmp) der „Brain Products GmbH“ verfügte über 32 Kanäle. Insgesamt wurden 34 Elektroden (32 Kanäle plus eine Referenz- und eine Erdungs-Elektrode) in einer Splitterbox zusammengeführt, siehe Abbildung 25. Über Elektrolytgel, das über einen Spalt in der Elektrodeneinhausung zwischen Elektrode und Kopfoberfläche des Probanden gespritzt werden konnte, sollten die Impedanzen minimiert werden. Die Abtastrate betrug 5000 Samples pro Sekunde pro Kanal. Die Splitterbox ihrerseits wurde über ein Flachbandkabel mit der sogenannten „actiCAP ControlBox“ verbunden. Die Spannungsversorgung der ControlBox erfolgte batterieelektrisch über sechs Volt. Über einen Signalverstärker, der seinerseits über die „BrainAmp PowerPack Box“ (Akku, 5.6 V) betrieben wurde, wurde die ControlBox an den EEG-Messrechner angeschlossen. Jeder einzelne EEG-Kanal besaß seine eigene elektronische Abschirmung („active shield“) zwischen Elektrodenende und Splitterbox. Auf diese Weise war es möglich, im Innenraum eines Fahrzeugs, in dem viele elektromagnetische Störeinflüsse vorherrschten, möglichst störungsfrei ein EEG abzuleiten.

Ebenfalls sorgte die Abschirmung für eine ausreichende Signalqualität, selbst wenn die Impedanzen zwischen Kopfoberfläche und Elektrode groß waren oder es zu Bewegungen der Kabel selbst kam. Diese Kabelbewegungen treten zwar selten im Labor, jedoch aufgrund der Vibrationen im fahrenden Pkw permanent auf.



Abbildung 25: EEG-„actiCAP“-System mit Impedanzwandlung.

Für die Messung von Alpha-Spindeln im EEG war lediglich die Verwendung von elf Kanälen vorgesehen (vgl. Simon et al., 2011). Simon et al. (2011) haben ursprünglich zwölf Kanäle verwendet. Der in Abschnitt 2.3.5.4 beschriebene Algorithmus zur Alpha-Spindeldetektion sieht keine Verrechnung zwischen einzelnen Kanälen vor – daher konnte die okzipital-zentrale Elektrodenposition (Oz) zum explorativen Schreiben eines EKG (nicht Gegenstand vorliegender Arbeit) verwendet werden. Die Elektrodenpositionen innerhalb der EEG-Hauben (bestehend aus elastischem Mischgewebe) wurden entsprechend des internationalen 10-20-Systems (Klem, Lüders, Jasper & Elger, 1999) vorbereitet, siehe Abbildung 26. Insgesamt standen EEG-Hauben für Kopfumfänge zwischen 54 und 60 cm zur Verfügung.

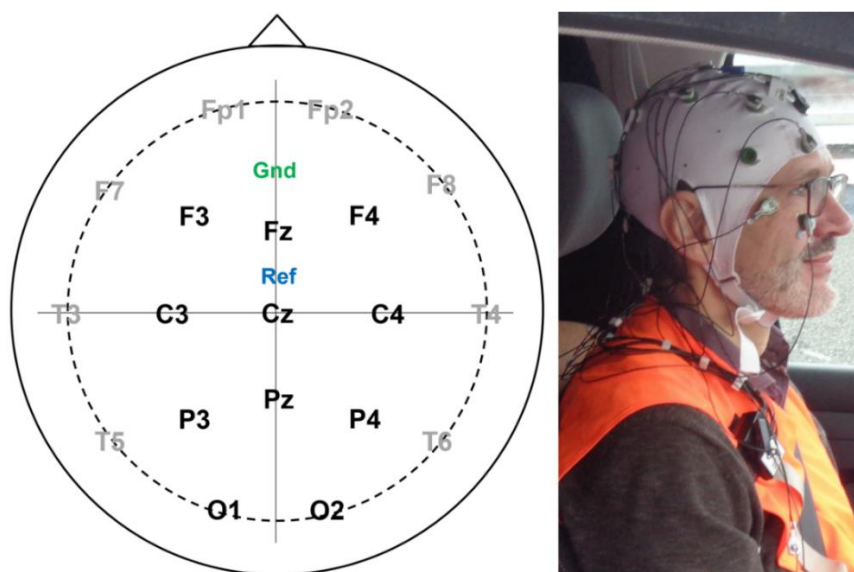


Abbildung 26: Links: Positionierung der EEG-Elektroden auf der Kopfoberfläche des Probanden entsprechend des „10-20“-Systems. Abkürzungen: Fp \triangleq frontal-parietal, F \triangleq frontal, C \triangleq zentral, P \triangleq parietal, O \triangleq okzipital, T \triangleq temporal, Gnd \triangleq Erdung, Ref \triangleq Referenz). Oben Nasion, unten Inion. Die fettgedruckten Elektrodenpositionen wurden im Experiment verwendet. Rechts: Proband mit EEG und EOG-Elektroden im „Wizard-of-Oz-Fahrzeug“.

Vier weitere Kanäle dienten der Aufzeichnung des EOG (Details zur Positionierung finden sich in Abschnitt 3.1.6). Das EOG diente dazu, Muskelartefakte, die aufgrund von Augenbewegungen entstanden, zu detektieren. Im Zuge der algorithmischen Alpha-Spindel-Detektion konnte so eine Artefakt-Rejektion durchgeführt werden. Um intraindividuelle Sprünge in der Alpha-Spindelrate zu glätten, wurden die Spindeln sowohl über einen gleitenden Mittelwert (vgl. Simon et al, 2011) von 60 Sekunden als auch über 300 Sekunden detektiert, und mit zehn Hertz aufgezeichnet. Für die Zeitreihenanalyse konnte so von der stärkeren Glättung über 300 Sekunden profitiert werden, wohingegen der gleitende Mittelwert von 60 Sekunden ereigniskorrelierte minutengenaue Analysen ermöglichte. Auch in dem in Abschnitt 2.4.5.2 vorgestellten Experiment von Schömig et al. (2015), in dem verschiedene Lidschlussindikatoren als Müdigkeitsmaß dienten, wurde sich für eine Glättung auf Basis eines gleitenden Mittelwerts über die Zeitreihen entschieden.

Alpha-Spindel-Analyse

Bei Auswertung der Alpha-Spindelrate jedes Probanden musste angelehnt an das von Schmidt (2010) bzw. Simon et al. (2011) beschriebene Vorgehen, eine probandenindividuelle Referenzierung vorgenommen werden. Von Simon et al. (2011) wurde ein Index berichtet, indem der Quotient aus der aktuellen Spindelrate und der Spindelrate aus der zuvor gefahrenen Baseline-

Fahrt bestimmt wurde. Zielsetzung der vorliegenden Arbeit war es jedoch, dass die um einen individuellen Startwert bereinigte Alpha-Spindelrate eines jeden Probanden sichtbar erhalten bleibt. Probandenindividuell wurden daher von der Alpha-Spindelrate zu einem Messzeitpunkt die jeweilige Alpha-Spindelrate zu Minute fünf subtrahiert: Wenn ein gleitender Mittelwert von 300 Sekunden, d.h. fünf Minuten für die Spindel detektion angelegt wurde, liegt erst nach fünf Minuten automatisierter Fahrdauer die erste „reine“ Alpha-Spindelrate vor, die ausschließlich während des automatisierten Fahrens (und beispielsweise nicht während des vorherigen Stillstands des Fahrzeugs) aufgezeichnet wurde. In der Konsequenz können alle Probanden auf eine Alpha-Spindelrate von null bei Minute fünf normiert und für jede Minute automatisierter Fahrt eine normierte (relative) Alpha-Spindelrate bestimmt werden.

Für die zufallskritische Analyse der Alpha-Spindelraten wurde sich für eine zeitliche Aggregation der Messwerte entschieden. Das zugrundeliegende experimentelle Design (vgl. Abschnitt 3.1.1) sah vor, dass in der Bedingung „niedrige Funktionszuverlässigkeit“ in jeder vierten gefahrenen Runde (d.h. ca. alle acht Minuten) ein Automatisierungsfehler präsentiert wurde. Demnach konnte für jeden Probanden individuell eine vergleichbare mittlere Alpha-Spindelrate über eine Fahrdauer von vier Runden gebildet werden, wobei zu berücksichtigen ist, dass der Automatisierungsfehler immer in der letzten Runde eines Aggregationsbereichs auftrat. Alle aggregierten Messwerte beinhalten demnach einen zeitlichen Anteil vor und einen zeitlichen Anteil nach dem Fahrfehler. Ein Beispiel: Der siebte Automatisierungsfehler in der niedrigen Zuverlässigkeitsbedingung trat ungefähr in der Mitte von Runde 28 auf, dieser befand sich in Aggregationsbereich sieben. In diesen Aggregationsbereich wurden ebenfalls die Messwerte aus Runde 25, 26 und 27 mit einbezogen. Das bedeutet, der Anteil im aggregierten Wert, der zeitlich gesehen *nach* dem Erleben des Automatisierungsfehlers einzuordnen ist, beträgt eine halb Runde (d.h. 12.5 %), da mit Beginn von Runde 29 bereits der achte Aggregationsbereich begann. Demzufolge sollten Auswirkungen in der Alpha-Spindelrate, die sich auf die Präsentation eines Automatisierungsfehlers im Aggregationsbereich (x) zeigen sollten, vorwiegend in dem nachfolgenden Aggregationsbereich (d.h. $x + 1$) interpretiert werden. Da auch hier die 33. Runde aus der Auswertung ausgeschlossen wird, werden insgesamt acht aggregierte Messwerte von jedem Probanden bis einschließlich der 32. durchfahrenen Runde berechnet.

Die analoge Vorgehensweise wird entsprechend auch für die hohe Zuverlässigkeitsbedingung angewendet, sodass der sechste von acht aggregierten Messwerten noch vollständig ohne Einfluss des nach ca. 50 Minuten auftretenden Automatisierungsfehlers ist, jedoch der siebte Messwert bereits einen zeitlichen Anteil der gemessenen Alpha-Spindelrate vor (ca. 87.5 %) und nach (ca. 12.5 %) dem einzigen erlebten Automatisierungsfehler aufweist.

Auch im Zuge dieser Aggregation musste eine Transformation der aufgezeichneten Alpha-Spindeln erfolgen, um alle individuellen Probandenraten vergleichbar zu normieren. Es wurde entschieden – kongruent zur Vorgehensweise der minütigen Alpha-Spindelrate – eine Differenz zum Startzeitpunkt zu bilden. Daher wird erneut die Alpha-Spindelrate zu Minute fünf gewählt, die von allen aggregierten Messzeitpunkten subtrahiert wird. Dies hat jedoch zur Folge, dass kein erster Aggregationsbereich (Runde eins bis vier) vorliegt, da die Minute fünf Bestandteil von diesem wäre, sondern nur der Startbereich „null“ als solcher als Bezugswert für die zufalls-kritischen Vergleiche existiert (hier jedoch als erster Aggregationsbereich benannt wird).

Alpha-Spindelraten-basierte Ermüdungsklassifikation

Um vermeintlich „ermüdete“ von „nicht ermüdeten“ Probanden auf Basis ihrer Alpha-Spindelraten zu trennen (UV 3), mussten entsprechende Kriterien definiert werden. Die Kriterien, die hierfür zugrunde gelegt wurden, bezogen sich auf das Zeitintervall von 40 Minuten Level 2-Fahrt, d.h. von Beginn des Versuchs (hier Minute fünf) bis zur Minute 45. Diese Begrenzung fußte auf dem experimentellen Design des Versuchs: Da die Probanden in der Bedingung „hohe Funktionszuverlässigkeit“ im Zeitbereich um Minute 50 einen einzigen Automatisierungsfehler erlebten, jedoch bei Probanden in der Bedingung „niedrige Funktionszuverlässigkeit“ alle acht Minuten ein Automatisierungsfehler auftrat, wäre der *gesamte* Zeitbereich (bis Minute 60) nicht konsistent vergleichbar gewesen.

Hypothesengeleitet (vgl. Abschnitt 2.5.1) wurde generell von einer Ermüdung aller Probanden über den Zeitbereich ausgegangen. Daher sollten Probanden identifiziert werden, die scheinbar nicht ermüdeten. Folgende Definition „nicht ermüdeten“ Probanden wurde aufgestellt und über eine Kombination zweier Hochpassfilter (a und b) umgesetzt: Wenn die maximale Alpha-Spindelrate nicht um mindestens drei Spindeln pro Minute zunahm (a) oder mehr als 25 % der minütigen Messwerte (d.h. in Summe 10 Minuten) unter dem Startwert null lagen (b), wurde der Proband als „nicht ermüdet“ klassifiziert. Bei der Auswahl der Filterkriterien wurde auf die Expertise der Autoren von Simon et al. (2011) zurückgegriffen, deren Erfahrung nach Probanden eine Alpha-Spindelrate von mindestens zwei Spindeln pro Minute in einem Messzeitraum aufweisen sollten, um generell Ermüdungsverhalten sichtbar machen zu können. Obwohl für Filter (a) ein konservativeres Kriterium von drei Spindeln statt zwei Spindeln pro Minute angesetzt wurde, sollte vermieden werden, dass – aufgrund einer hohen Inkonsistenz in den Zeitreihen – eine Vielzahl von Probanden als „ermüdet“ eingestuft werden, die lange Zeitbereiche der Versuchsfahrt geringere Alpha-Spindelraten als zu Beginn der Messung (d.h. negative Al-

pha-Spindelraten) aufweisen. Um dies zu kontrollieren, wurde ein zusätzlicher Filter (b) gewählt, der jedoch nicht zu konservativ sein durfte, wenn manche Probanden erst mehrere Minuten nach Messzeitpunkt fünf einen Anstieg der Alpha-Spindelrate aufwiesen. Außerdem war für diesen Filter (b) eine Robustheit gegenüber Schwankungen in der Alpha-Spindelrate um den Wert null wichtig. Daher wurden Anteile der Messwerte (25 %) über einen Zeitraum definiert, statt lediglich den Zeitreihenbeginn zu berücksichtigen. Eine noch konservativere Auslegung beider Filter sollte vermieden werden, um für zufallskritische Analysen eine angemessene Stichprobengröße der vermeintlich „ermüdeten“ Probanden vorhalten zu können.

3.1.4.4 Fragebögen

Auf einem demographischen Fragebogen wurde der Proband gebeten, persönliche Angaben zu vermerken, die zur Beschreibung der Stichprobe dienlich sind. Hierunter fiel das Alter, das Geschlecht, der höchste Schul- oder Bildungsabschluss, das aktuelle Beschäftigungsverhältnis, das Jahr der Pkw-Fahrerlaubnisprüfung, die jährliche Fahrleistung und die Abfrage, ob die Fahrtüchtigkeit gegenwärtig durch Medikamenteneinnahmen beeinträchtigt war. Falls der Proband angegeben hätte, dass seine Fahrtüchtigkeit beeinträchtigt gewesen sei, wäre der Fahrversuch nicht durchgeführt worden.

Um subjektive Daten zu erheben, wurden zwei Fragebogenhefte für die Vor- und Nachbefragung zusammengestellt. Diese befinden sich in Anhang A und B. Das Heft für die Vorbefragung (Teil 1) umfasste zunächst die grundlegende Information, was genau unter teilautomatisiertem Fahren zu verstehen ist. Es wurde aus Gründen der Informationssparsamkeit bewusst auf den Zusatz „Level 2“ verzichtet, da die Kenntnis über die Einteilung von kontinuierlich automatisierten Fahrfunktionen in verschiedene Level für diesen Versuch keinen Mehrwert für den Probanden bedeutet hätte. Unmittelbar nach dieser Erklärung sollte der Proband einzig auf Basis der Funktionsbeschreibung (im Fragebogen „Systembeschreibung“) auf einer siebenstufigen Likert-Skala einschätzen, ob er im Alltag eine teilautomatisierte Fahrfunktion nutzen wollen würde. Ebenfalls war von Interesse, ob der Proband unter Umständen Vorerfahrung mit bestimmten Fahrerassistenzsystemen besaß. Hierfür wurden zehn häufig in Fahrzeugen verbaute Assistenzsysteme ausgewählt (vgl. Gruschwitz et al., 2020), deren regelmäßige Nutzung letztlich dichotom beantwortet werden konnte (der Zusatz „bin mir nicht sicher“ stand zwar auch zur Auswahl für den Probanden, sollte jedoch in der Auswertung äquivalent zu einem „nein“ behandelt werden).

Sehr früh im Fragebogenheft (bereits auf der zweiten von sieben Seiten) sollte der Proband seine aktuelle Schläfrigkeit bewerten. Diese frühe Abfrage sollte vermeiden, dass womöglich die Bearbeitung des Fragebogenheftes selbst zu leichten Ermüdungseffekten führt, die die primäre Erhebung konfundieren könnten. Die verwendete Skala entstammte dem subjektiven Teil der „Subjektiven und objektiven Müdigkeitsskala von Braunschweig“ (SOMS) aus Reinprecht (2013), vgl. Abschnitt 2.3.5.1. Diese Abfrage wiederholte sich in der Nachbefragung (Teil 2), sodass ein unmittelbarer Vergleich der subjektiv beurteilten Müdigkeit zwischen der Vor- und Nachbefragung möglich war.

Hieran schloss sich der „Motion Sickness Susceptibility Questionnaire“ (MSSQ; Golding, 2006) an, der beansprucht, mittels neun Items zu Kinetoseerfahrungen in der Kindheit (vor dem zwölften Lebensjahr) und neun Items zu Kinetoseerfahrungen in den letzten zehn Jahren, die Anfälligkeit für Kinetose zu prädictieren. Alle Items beziehen sich dabei auf das Auftreten von kinetostypischen Symptomen, wie Übelkeit zu erleben oder Erbrechen zu müssen, wenn Verkehrsmittel oder andere Fahrzeuge (z.B. Karusselle oder Achterbahnen) genutzt werden. Golding (2006) hebt die Gütekriterien dieses Testinstruments hervor: Er berichtet in Bezug auf die Reliabilität von einem Cronbach's Alpha von .87 und einer Test-Retest-Reliabilität von ca. $r = .90$. Die prädiktive Validität von $r = .51$ wurde auf Basis von Bewegungsexperimenten im Labor erhoben und liegt damit nach Golding (2006) höher als konstruktäquivalente Testinstrumente aufweisen. Alle Items wurden für eine deutschsprachige Stichprobe aus dem Englischen ins Deutsche übersetzt. Für die Berechnung der Punktskoren wurde die von Golding (2006) vorgeschlagene Norm (vgl. Abschnitt 3.2.2.6) zugrunde gelegt.

Mittels der „Automation Trust Scale“ nach Jian, Bisantz und Drury (2000) wurde das Vertrauen in eine teilautomatisierte Fahrfunktion – wieder ausschließlich auf Basis der dem Probanden zur Verfügung stehenden Funktionsbeschreibung – erhoben. Demnach musste sich der Proband vorstellen, wie zuverlässig die Funktion wohl arbeiten wird und sein Vertrauen hieraufhin einschätzen. Auch diese methodische Vorgehensweise ist vor dem Hintergrund der analogen Abfrage in der Nachbefragung zu sehen: Indem nach dem tatsächlichen Erleben des teilautomatisierten Fahrens die „Automation Trust Scale“ erneut eingesetzt wurde, konnte der Effekt der Nutzungserfahrung in Hinblick auf das Vertrauen in die Automatisierungsfunktion bestimmt werden. Insgesamt besitzt die Skala zwölf Items (jeweils sieben Kategorien). Um das inhaltliche Verständnis der deutschsprachigen Probanden reliabel zu gewährleisten, wurden die Items von Englisch in Deutsch übersetzt. Hierbei wurde größtenteils auf bereits angewandte Übersetzungen aus Beggiato (2015) oder aus Vogelpohl, Vollrath, Kühn, Hummel und Gehlert (2016) zurückgegriffen.

Auf Basis der Grundlagenforschung ist davon auszugehen, dass die Bearbeitung von Vigilanzaufgaben mit einer Erhöhung des subjektiv empfundenen Stresses einhergeht (vgl. Abschnitt 2.2.3.2). Um das subjektive Stressniveau des Probanden vor und nach der Versuchsfahrt zu erheben, wurde der „Short Stress State Questionnaire“ (SSSQ) von Helton (2004) ausgewählt. Helton (2004) berichtet, dass eine Itemreduktion der ursprünglichen 90 Items des „Dundee Stress State Questionnaire“ (DSSQ) auf nun insgesamt 24 Items vorgenommen wurde. Dabei wurden die drei zugrundeliegenden Dimensionen („task engagement“, „distress“ und „worry“) beibehalten, wobei Wert daraufgelegt wurde, dass die finalen 24 Items sehr trennscharf sind, keine inhaltlichen Dopplungen aufweisen und eine adäquate Abdeckung der Faktorenstruktur gewährleisten. Hiervon ausgehend berichtet Helton (2004) ein Cronbach's Alpha zwischen .81 und .87 und führte eine erfolgreiche Validierung anhand einer Vigilanzaufgabe im Labor durch. Für die vorliegende Studie mussten alle Items zunächst ins Deutsche übersetzt werden. Bereits hier zeigte sich, dass die ursprüngliche, feine sprachliche Differenzierung der Aussagen im Englischen nach der Übersetzung ins Deutsche zu uneindeutigen, inhaltlichen Dopplungen führen, die Helton (2004) explizit vermeiden wollte. Insbesondere die Dimensionen „task engagement“ (Item 11) und „worry“ (Items 16 und 18) sind hiervon betroffen. Es wurde entschieden – in Ermangelung eines alternativen, ebenso effizienten Verfahrens zur Abschätzung subjektiv empfundenen Stresses – nur 21 der 24 Items des SSSQ in der vorliegenden Studie zu verwenden. Demnach muss berücksichtigt werden, dass sich die psychometrischen Eigenschaften des Testverfahrens nicht mehr auf den angepassten Test übertragen lassen. Es wird an dieser Stelle angenommen, dass die vorhandenen, semantisch sinnvollen Items subjektiv empfundenen Stress dennoch verhältnismäßig valide abbilden.

Wesentliche Zielsetzung vorliegender Untersuchungen war, dass die erzielten Ergebnisse mit hoher Wahrscheinlichkeit auf die zugrundeliegenden experimentellen Manipulationen zurück zu führen sind. Um bei Bedarf Konfundierungen durch differenzielle psychologische Merkmale ausschließen zu können, wurde ein grundlegender Test der prominentesten Persönlichkeitsmerkmale, den „Big Five“, entsprechend in der Vorbefragung integriert. Damit die Probanden nicht länger als nötig der Fragebogenbearbeitung ausgesetzt waren, wurde sich auch hier für einen Kurztest entschieden. Der verwendete „BFI-10“ nach Rammstedt und John (2007) beansprucht effizient mittels zehn Items die Konstrukte „Extraversion“, „Neurotizismus“, „Gewissenhaftigkeit“, „Verträglichkeit“ und „Offenheit für neue Erfahrung“ anhand von jeweils zwei Items pro Konstrukt zu erfassen. Ein elftes Item kann optional auch zur Erfassung der „Verträglichkeit“ herangezogen werden. Alle Items wurden auf einer fünfstufigen Likert-Skala beantwortet. Hinsichtlich der konvergenten Validität zum NEO-Persönlichkeitsinventar (NEO-

PI-R) berichten Rammstedt und John (2007) eine Korrelation von .48, wobei insgesamt im Vergleich zur Langversion des BFI (dem „BFI-44“) Einbußen in der Varianzaufklärung von acht Prozent in Kauf genommen werden müssen.

Im Anschluss an den Fahrversuch wurde dem Probanden das Fragebogenheft (Teil 2) ausgehändigt. Über ein einzelnes Item (zehnstufige Likert-Skala) wurde das aktuelle Befinden des Probanden abgefragt. Dieses Item stellte die Nachbefragung anhand des zuvor beschriebenen MSSQ dar und verwendete die Verbalanker „Ich fühle mich gut“ und „Ich fühle mich furchtbar, als ob ich mich übergeben müsste“. Ebenfalls war es unmittelbar im Anschluss an den Fahrversuch wichtig, die subjektive Müdigkeit wiederholt mit dem subjektiven Teil der SOMS zu messen, um einen Vergleich zur Vorbefragung ziehen zu können.

Passend zur Cover-Story wurde der Proband nach seiner Zufriedenheit mit der automatisierten Fahrfunktion gefragt und wurde gebeten, sein Nutzungsinteresse (jeweils auf einer neunstufigen Likert-Skala) anzugeben. Eine offene Frage, die auf Verbesserungsmöglichkeiten abzielte, sollte dem Probanden Raum geben, (unbewusst) positive wie negative Aspekte des eingesetzten Wizard-of-Oz-Konzepts zu äußern.

Als weiteres Werkzeug zur Erfassung der Beanspruchung durch die Vigilanzaufgabe wurde zusätzlich zum SSSQ das prominente Testinstrument „NASA-Task Load Index“ (NASA-TLX; Hart & Staveland, 1988) zur Abschätzung des benötigten Workloads in Bezug auf die Vigilanzaufgabe eingesetzt (vgl. auch Abschnitt 2.2.3.2). Die Test-Retest-Reliabilität liegt nach Hart und Staveland (1988) bei .83. Es wurden alle sechs Subskalen (jeweils zehnstufig likert-skaliert) „mentale (bzw. im Fragebogenheft „geistige“) Anforderung“, „körperliche Anforderung“, „zeitliche Anforderung“, „Aufgabenerfüllung“, „Anstrengung“ und „Frustration“ verwendet. Hierdurch war sowohl eine ganzheitliche Workload-Analyse über alle Skalen als auch über die Subskalen möglich. In der Nachbefragung kamen ebenfalls der SSSQ und die „Automation Trust Scale“ zum Einsatz, um einerseits den subjektiv eingeschätzten Stress und andererseits das Vertrauen in die automatisierte Fahrfunktion mit der Vorbefragung vergleichen zu können.

Auch das Komforterleben des Probanden sollte in Abhängigkeit von seiner Ermüdung und der Funktionszuverlässigkeit – nachdem der Proband eine Level 2-Fahrt erlebte – gemessen werden. Hierfür wurde das Testinstrument „Discomfort in Automated Driving“ („The Discomfort Scale“; Siebert, Oehl, Höger & Pfister, 2013) herangezogen. Siebert et al. (2013) berichten hin-

sichtlich der Reliabilität ein Cronbachs's Alpha von .91 als Maß der internen Konsistenz. Analog zu Vogelpohl et al. (2016) wurde die Disco-Scale für die vorliegende Untersuchung jedoch leicht modifiziert und gekürzt, sodass die von Vogelpohl et al. (2016) vorgeschlagenen deutschsprachigen Übersetzungen sinnvoll angewendet werden konnten. Auch hier sollte dementsprechend berücksichtigt werden, dass die von Siebert et al. (2013) berichteten psychometrischen Eigenschaften der Skala sehr wahrscheinlich verletzt sind. Insgesamt zehn der ursprünglichen 14 Items wurden verwendet, wobei jedes Item auf einer siebenstufigen Likert-Skala beantwortet werden sollte.

Letztendlich sollte im Rahmen der Studie auch ein Augenmerk auf die Immersions-Güte der Wizard-of-Oz-Methodik geworfen werden. Es war von besonderem Interesse, ob der Proband tatsächlich glaubte, von einer automatisierten Fahrfunktion und nicht von einem Menschen gefahren zu werden. Falls dem Probanden nach der Studie die Frage gestellt worden wäre „Wussten Sie, dass Sie von einem menschlichen Fahrer gefahren wurden?“, hätte der Proband sehr wahrscheinlich auf diese Suggestion mit einem „Ja“ geantwortet (auch wenn er es vielleicht gar nicht erahnt hatte). Daher wurde eine offene Frage verwendet, die indirekt auf die Immersionsgüte des Wizard-of-Oz-Ansatzes rückschließen lässt: „Auf welche Art und Weise wurde Ihrer Einschätzung nach das teilautomatisierte Fahren in diesem Versuch technisch umgesetzt?“. Wenn der Proband hier beispielsweise verschiedene Sensortechniken benannt hatte, wurde davon ausgegangen, dass er nichts von den Steuerungstätigkeiten des Versuchsleiters (d.h. des Wizard-Fahrers) erahnte. Dennoch hätte die Frage dem Probanden auch die Möglichkeit gegeben, hier die Involviertheit eines menschlichen Fahrers zu mutmaßen.

Da insbesondere die Studie von Vogelpohl et al. (2017) zeigte, dass die Schlafdauer vor dem Fahrversuch einen zu berücksichtigenden Faktor für die Müdigkeitsentwicklung darstellt, wurde entschieden, diese zwar nicht zu kontrollieren, jedoch zu erheben. Ebenso verhielt es sich in Bezug auf den Koffeinkonsum (vgl. Smith, 2002). Es interessierte, ob die zuvor genannten Faktoren womöglich Einfluss auf die Ermüdung des Probanden genommen haben. Die Nachbefragung wurde folglich mit einer Angabe über die Schlafdauer in der letzten Nacht als auch mit Angabe zum vorherigen Koffeinkonsum (dichotom) abgeschlossen.

3.1.5 Instruktion

Die Probanden wurden schriftlich über ihre Aufgabe während des Versuchs instruiert. Eingangs wurde grob erläutert, dass es Zielsetzung der Studie sei, Auswirkungen von teilautomatisiertem Fahren auf den Fahrer zu untersuchen. Weiter hieß es im Text, dass das EEG zwecks Aufmerksamkeitsmessung verwendet wird. Explizit wurde nicht erläutert, dass entsprechend der Forschungshypothesen von erhöhter Ermüdung und einer abnehmenden Aufmerksamkeitsleistung des Fahrers über die Fahrdauer ausgegangen wird. Auch über die genaue Dauer der automatisierten Fahrt wurden die Probanden im Unklaren gehalten, es wurde lediglich angegeben, dass die Fahrt weniger als 90 Minuten dauern wird. Hierdurch sollte möglichen Effekten durch Probandenkenntnis über das Ende des Versuchs (beispielsweise eigenes Leistungsmanagement) vorgebeugt werden.

Es wurde beschrieben, dass das automatisierte Fahrzeug das Oval mit einer (leicht variablen) Geschwindigkeit zwischen 60 und 80 km/h auf dem mittleren Fahrstreifen durchfährt. Um die kontrollierte Versuchsbedingung auf der Teststrecke hinsichtlich der externen Validität etwas zu erhöhen, sollten sich die Probanden vorstellen, auf einer Autobahn mit zwei Fahrstreifen und einem Seitenstreifen zu fahren. Demnach würde der mittlere Fahrstreifen der Teststrecke dem rechten Fahrstreifen auf einer Autobahn entsprechen.

Es wurde ferner erläutert, dass der Versuch durchgeführt wird, um eine Funktionsbewertung seitens des Probanden zu ermöglichen. Dieser Teilaspekt der Instruktion gehörte zur Cover-story, die bereits konsistent bei Probandenakquise verwendet wurde. Auf diese Weise sollte sich für die Probanden ein schlüssiges Bild einer Qualitätsbewertungsaufgabe ergeben: Den Versuchsteilnehmern wurde erklärt, dass – obwohl das Automatisierungssystem schon sehr weit entwickelt sei – gelegentlich Fahrfehler auftreten können. Mit Hilfe der Rückmeldung des Probanden über die Fehler sollte das System weiter verbessert werden. Mögliche auftretende Fehler wurden dabei konkret benannt: Eine Verringerung der Fahrgeschwindigkeit unter 60 km/h (Typ a), eine Erhöhung der Fahrgeschwindigkeit über 80 km/h (Typ b), ein Verlassen des Fahrstreifens nach links (Typ c) und ein Verlassen des Fahrstreifens nach rechts (Typ d).

Die Herausforderung im Rahmen der Wizard-of-Oz-Methodik bestand darin, eine kontinuierliche automatisierte Fahrt gemäß SAE Level 2 nachzubilden, mit dem Zusatz, dass zu keiner Zeit das Lenkrad berührt werden soll, also „hands-off“ gefahren wird. Somit wurde in der Instruktion erläutert, dass das „System“ so programmiert sei, dass die auftretenden Fehler wieder

selbstständig korrigiert werden würden und eine Fahrzeugsteuerungsübernahme nicht notwendig sei. Falls einer der zuvor beschriebenen Fehler erkannt werden würde, sollte der Proband einmal so schnell wie möglich den Handtaster drücken.

Auch für eine manuelle Fahrt (Baselinefahrt) vor Beginn der primären automatisierten Fahrt, wurde der Proband instruiert. Hierfür diente eine Abfolge von Fahraufgaben, die – gut lesbar in großer Schrift gedruckt – auf der Mittelkonsole des Versuchsfahrzeugs befestigt wurde. Alle neun Runden, die für die Baselinefahrt benötigt wurden, wurden hier definiert. Insofern war gewährleistet, dass der Proband ohne Stopp und ohne weitere Nachfragen die komplette Baselinefahrt absolvieren konnte. Die Abfolge der Fahraufgaben findet sich in Anhang E.

3.1.6 Versuchsdurchführung

Das Experiment fand im Zeitraum von Oktober bis Dezember 2016 statt. Die Nacherhebung von acht Probanden wurde ein Jahr später unter vergleichbaren jahreszeitlichen Witterungs- und Beleuchtungsbedingungen im Dezember 2017 durchgeführt. Bei 17 der 35 Versuchsfahrten herrschten trockene Bedingungen (kein Niederschlag, trockene Fahrbahn) vor. Hiervon entfallen auf die Bedingung „hohe Funktionszuverlässigkeit“ neun Fahrten und auf die Bedingung „niedrige Funktionszuverlässigkeit“ acht Fahrten. Nasse Versuchsbedingungen, d.h. Niederschlag und / oder eine nasse Fahrbahn wurden von den Probanden in 18 der 35 Versuchsfahrten erlebt (hohe Zuverlässigkeit: neun Fahrten, niedrige Zuverlässigkeit: neun Fahrten). Die Außentemperatur lag bei den Fahrten mit hoher Funktionszuverlässigkeit zwischen zwei und 13 °C ($M = 8.89$, $SE = .76$) und mit niedriger Funktionszuverlässigkeit zwischen zwei und 10 °C ($M = 6.06$, $SE = .65$). Der Innenraum des Fahrzeugs wurde auf ca. 21 °C beheizt.

Jeder Proband traf an seinem Versuchstag gegen 10:30 Uhr an der Teststrecke ein. Nach der Begrüßung durch den Versuchsleiter wurde der Teilnehmer auf direktem Weg zum Büroraum begleitet, ohne die Möglichkeit zu haben, das Wizard-of-Oz-Fahrzeug zu sehen. Im Büroraum wurde den Probanden in Abwesenheit des Versuchsleiters zunächst ein 15-minütiger Einweisungsfilm des ATC vorgeführt, um eine Einweisung in das Sicherheitskonzept auf dem Versuchsgelände zu gewährleisten. Im Anschluss hieran gab der Versuchsleiter eine grobe Übersicht über den Versuchsablauf (Ausfüllen von Einverständnis- und Datenschutzerklärung, Vorbefragung mittels Fragebögen, Anlegen des EEG, Instruktionen zum Versuch, Einweisung in das Fahrzeug, Durchführung des Fahrversuchs, Nachbefragung mittels Fragebögen, Ablegen des EEG und Waschen der Haare, Vergütung und Verabschiedung). Generell bestand für den

Probanden vor und während der Studie (außer bei EEG-Aufzeichnung) die Möglichkeit, beliebig viele Fragen zu stellen. Falls eine Beantwortung aus methodischen Gründen zum gegenwärtigen Zeitpunkt nicht möglich war, wurde höflich auf einen späteren Zeitpunkt für die Beantwortung verwiesen. Zu jedem Versuchszeitpunkt war es dem Probanden möglich, seine Teilnahme ohne die Angabe von Gründen abzubrechen. Er hätte dennoch seine Aufwandsentschädigung im vollen Umfang erhalten.

Im Büroraum wurde der Proband zunächst gebeten, die Einverständnis- und Datenschutzerklärung zu unterzeichnen. Um die richtige EEG-Kappengröße auszuwählen, bestimmte der Versuchsleiter den Kopfumfang des Probanden mit einem Maßband. So konnte der Versuchsleiter – während der Versuchsteilnehmer das erste Fragebogenheft (vgl. Anhang A) ausfüllte – die individuelle EEG-Kappe vorbereiten. Hierzu wurden alle Elektroden in den definierten Kappenpositionen befestigt sowie die Elektroden zum Aufzeichnen des EOG und des EKG mit Kleberingen versehen. Sobald der Proband mit der Bearbeitung des Vorabfragebogens fertig war, konnte er noch einmal Mineralwasser trinken bzw. die Toilette aufsuchen. Der Versuchsteilnehmer zog zunächst eine Warnweste über seine Kleidung. Diese hatte eine kombinierte Funktion: Einerseits diente die Weste als mobile Befestigung für die Splitterbox, indem eine Zugentlastung auf der Schulter befestigt werden konnte (mittels Klettband) und die ActiCAP ControlBox in einer aufgenähten Seitentasche der Weste verstaut wurde. Andererseits wurden so die Sicherheitsbestimmungen auf dem Testgelände erfüllt, falls der Proband unerwarteterweise das Versuchsfahrzeug verlassen müsste. Dem Versuchsteilnehmer konnte nun die EEG-Kappe aufgesetzt werden. Hierzu richtete der Versuchsleiter die Kappe so aus, dass sie zentriert auf der Kopfoberfläche auflag; war die optimale Position gefunden, wurde die Kappe mit Hilfe eines Kinnbandes fixiert. Im Anschluss wurden drei Elektroden zum Aufzeichnen des EOG auf der Haut um die Augen befestigt (EOG-rechts, EOG-links, EOG-unten (unter dem rechten Auge)). Eine vierte Elektrode zur EOG-Aufzeichnung (EOG-oben (über dem rechten Auge)) war bereits im unteren Rand der EEG-Kappe integriert. Die EKG-Elektrode klebte sich der Proband meist selbst unter Anleitung des Versuchsleiters in einem Abstand von fünf Zentimetern auf Höhe des Brustbeins nach links. Nun konnten alle Elektroden mit Elektrolytgel, das in eine Spritze aufgezogen wurde, gefüllt werden. Die aktiven Elektroden zeigten hierfür über drei Leuchtfarben (rot, gelb und grün) an, ob die Impedanzen zur Kopfoberfläche gering genug waren (d.h. im grünen Bereich). Sobald alle Elektroden hinreichend gefüllt waren, war das Anlegen des EEGs abgeschlossen. Der Versuchsleiter handigte dem Probanden nun die Instruktion für den Fahrversuch aus (vgl. Abschnitt 3.1.5). Während der Proband die Instruktion las, traf der Versuchsleiter die letzten Vorbereitungen am Wizard-of-Oz-Fahrzeug. Hierzu zählten das Starten des Motors, die Aktivierung der zweiten Fahrzeugsteuerungseinrichtung im Fond und

das Starten des Tablett-PC. Auf dem Tablett-PC wurde eine teilweise permutierte Reihenfolge (die Permutationsrestriktionen sind in Abschnitt 3.1.1 erläutert) randomisiert für den jeweiligen Probanden geladen.

Der Versuchsleiter erkundigte sich beim Probanden, ob dieser die Instruktionen vollständig verstanden habe, oder noch Rückfragen bestünden. Bei Rückfragen sorgte der Versuchsleiter dafür, lediglich mit anderen Worten ein besseres Verständnis herbeizuführen, anstatt mehr Informationen preiszugeben. Weiterhin wurde dem Probanden versichert, dass zum entsprechenden Zeitpunkt im Versuch die Instruktion mündlich wiederholt werden wird. Der Proband wurde nun vom Versuchsleiter auf schnellstem Wege zum Wizard-of-Oz-Fahrzeug geführt und gebeten, auf dem Fahrersitz Platz zu nehmen. Es bestand genügend Zeit, um eine individuelle Sitz- und Außenspiegeleinstellung vorzunehmen. Da viele Probanden nicht regelmäßig ein Fahrzeug mit Automatikgetriebe fuhren, wurde ihnen die Bedienung erläutert. Ferner wurde der Proband mit Sicherheitsmerkmalen des Wizard-of-Oz-Fahrzeugs vertraut gemacht; über lautes Sprechen konnte sich der Proband jederzeit – auch durch die Barriere der Trennscheibe – bemerkbar machen. Anschließend wurde das EEG an den Messrechner angeschlossen und der vorbereitete Handtaster mittels Kreppband an einem beliebigen Finger der rechten Hand des Probanden befestigt. Auf diese Weise konnte sichergestellt werden, dass der kleine Handtaster nicht versehentlich während der Versuchsfahrt in den Fahrzeugfußraum fiel. Eine sicherheitskritische Einschränkung der Fahrzeugsteuerung durch die Verkabelung konnte ausgeschlossen werden. Nun ging der Versuchsleiter den Ablauf des gesamten Versuchs mit dem Probanden zusammen durch: Fahrt auf die Ovalbahn der Teststecke, EOG-Baselinemessung im Stand, manuelle Fahrt über neun Runden, „Kalibrierungsfahrt“ der Automatisierungsfunktion über zwei Runden (Erklärung folgt) und die primäre Versuchsfahrt, die entsprechend der schriftlichen (vorab gelesenen) Instruktionen verlaufen sollte. Bestanden keine Rückfragen seitens des Probanden mehr, wurde der EEG-Messrechner auf dem Beifahrerplatz des Versuchsfahrzeugs gestartet und der Versuchsleiter nahm den zweiten Fahrerplatz im verdunkelten Fond ein.

Über ein Funkgerät konnte der Versuchsleiter mit dem Dispatcher an der Teststrecke kommunizieren. Sobald dieser die Freigabe zur Fahrt auf die Ovalbahn gab (wenn kein anderes Fahrzeug mehr die Ovalbahn befuhr), gab der Versuchsleiter seinerseits die Aufforderung an den Probanden, bitte die Teststrecke zu befahren. Der Proband fuhr mit ca. 70 km/h etwas weniger als eine Runde auf dem mittleren Fahrstreifen durch das Oval. Währenddessen prüfte der Versuchsleiter, ob alle Signale der Messtechnik (d.h. Spannungsverläufe der EEG-Einzelelektroden, Videodaten der vier Kameras, Eingaben des Probanden über den Handtaster sowie Daten

der Fahrgeschwindigkeit und Spurposition von der Sensorkamera) zur Verfügung standen. An einer definierten Position („Start“-Markierung in Abbildung 19; Abschnitt 3.1.4.1) wurde der Proband gebeten, anzuhalten. Diese Position befand sich kurz hinter der zweiten Steilwandkurve auf dem ersten geraden Teilstück der Strecke.

Zunächst wurde im Stillstand des Fahrzeugs vom Versuchsleiter die Eingabe der Probandennummer und Versuchsbedingung (A oder B) vorgenommen und die Datenaufzeichnung gestartet. Unmittelbar darauf startete die EOG-Baselinemessung für eine Dauer von etwa drei Minuten. Der Proband wurde durch eine eingesprochene (weibliche) Stimme (präsentiert über die Lautsprecher des EEG-Messrechners) aufgefordert, verschiedene Augenbewegungen durchzuführen. Folgende Kommandos wurden in dieser Reihenfolge ausgegeben: „Augen nach links“, „Augen nach rechts“, „Augen nach oben“, „Augen nach unten“, „Augen rollen links“, „Augen rollen rechts“, „Augen zusammenkneifen“, „bitte zehn Sekunden mit den Augen blinzeln“, „Augen schließen und entspannen“. Nach Beendigung einer jeden angesagten Augenbewegung folgte das Kommando „zurück“, woraufhin der Proband wieder geradeaus durch die Windschutzscheibe blicken sollte. Mit Hilfe dieser standardisierten Abfolge von Augenbewegungskommandos wurde der für die Alpha-Spindelidentifikation zuständige Algorithmus probandenindividuell für die Artefakterkennung im EEG angelernt, um während der EEG-Aufzeichnung eine Artefakt-Rejektion durchführen zu können.

Sobald die EOG-Baselinemessung beendet war, wurde der Proband gebeten, entsprechend der Anweisung auf der Mittelkonsole (vgl. Anhang E) neun Runden manuell durch das Oval zu fahren. Die entsprechenden Aufgaben während der Fahrt (Fahrgeschwindigkeitsänderungen und Fahrstreifenwechsel) sollten typischen Fahrmanövern im öffentlichen Straßenverkehr nachempfunden sein. Während dieser Fahrdauer wurde eine EEG-Baselinemessung durchgeführt. Hierbei trainierte der Algorithmus die Erkennung von Alpha-Spindeln probandenindividuell über eine Dauer von 15 Minuten Fahrzeit, nachdem fünf Minuten Einführungsfahrt verstrichen waren.

Im Anschluss an die EEG-Baselinemessung brachte der Proband das Wizard-of-Oz-Fahrzeug auf Bitte des Versuchsleiters zum Stillstand (an der „Start“-Position, an der auch die EOG-Baselinemessung stattfand). Nun schaltete der Versuchsleiter auf die zweite Steuerungseinrichtung um. Dem Probanden wurde kurz in Erinnerung gerufen, dass nun eine „Kalibrierungsfahrt“ der automatisierten Fahrfunktion stattfände. Tatsächlich diente die automatisierte Fahrt über zwei Runden dazu, dass dem Probanden erstmalig automatisiertes Fahren vorgestellt wurde

(um Effekte erstmaliger Exposition abzumildern) und der Versuchsleiter die Gelegenheit bekam, sich in das möglichst perfekte Durchfahren der Ovalbahn einzufinden. Der Versuchsleiter beschleunigte hierzu das Fahrzeug auf 70 km/h im mittleren Fahrstreifen. Der Proband seinerseits konnte die Fahrt beobachten, ohne dass ihm eine spezifische Aufgabe gestellt wurde.

Nach der „Kalibrierungsfahrt“ hielt der Versuchsleiter das Fahrzeug wieder an der markierten Position auf der Ovalbahn an. Dem Probanden wurde kurz erläutert, dass nun der primäre Versuch beginne: Der Proband sollte permanent die Geschwindigkeit und Spurhaltung des Versuchsfahrzeugs überwachen und immer dann schnellstmöglich den Handtaster drücken, sobald er Abweichungen vom Soll feststelle. Auch hier beschleunigte der Versuchsleiter das Fahrzeug auf 70 km/h im mittleren Fahrstreifen. In dieser ersten Studie wurde keine Geschwindigkeitsregelanlage verwendet; auf diese Weise traten immerzu leichte Änderungen der Fahrgeschwindigkeit (innerhalb des Solls) auf. Der Versuchsleiter durchfuhr 33 Runden während er nach jeder absolvierten Runde auf dem Tablett-PC eine Anzeige weiter schaltete (vgl. Abbildung 22; Abschnitt 3.1.4.2). Zusätzlich wurde dem Versuchsleiter auf dem Bildschirm des Messrechners die aktuelle Runde angezeigt, die über GPS ermittelt wurde: Indem vier GPS-Marker auf der Ovalbahn nacheinander passiert wurden, konnte jeweils eine neue Runde zuverlässig gezählt werden.

Sobald alle Runden absolviert waren, beendete der Versuchsleiter die Fahrt, indem er an dem markierten Standort auf der Ovalbahn das Fahrzeug anhielt. Unmittelbar darauf wurde die Datenaufzeichnung gestoppt und alle Messdaten in einer Gesamtdatei gespeichert. Der Proband wurde nun instruiert, die Teststrecke zu verlassen. Nachdem alle Steckverbindungen vom EEG zum EEG-Messlaptop getrennt wurden, nahm der Proband erneut im Büroraum Platz. Hier bearbeitete er sofort das zweite Fragebogenheft (vgl. Anhang B) in Abwesenheit des Versuchsleiters. Dieser fuhr zwischenzeitlich die zuvor aktivierten Komponenten im Versuchsfahrzeug herunter und erstellte eine Sicherungskopie der Messdaten. Nach vollständiger Bearbeitung des zweiten Fragebogens wurde dem Probanden das EEG wieder abgelegt und er konnte seine Haare in einem Duschaum waschen. Schließlich erläuterte der Versuchsleiter den fachlichen Hintergrund und klärte über die Täuschungen durch die Cover-Story und des Wizard-of-Oz-Fahrzeugkonzepts auf. Der Proband erhielt für die Versuchsdauer von ca. vier Stunden eine Aufwandsentschädigung in Höhe von 80 Euro und wurde verabschiedet. Nach dem Einsatz des EEG-Systems wurde die verwendete Haube mit Wasser und mildem Haarshampoo gewaschen und alle einzelnen Elektroden von Elektrolytgel-Rückständen befreit.

3.1.7 Videobasiertes Müdigkeitsrating

Während der Versuchsfahrt wurde der Proband mittels Videokameras gefilmt (vgl. Abschnitt 3.1.4.2). Dieses Videomaterial sollte im Anschluss an den Fahrversuch hinsichtlich der Müdigkeit des Probanden von einer geschulten Person („Rater“) beurteilt werden. Die Software „Mued 8.1 – Macht Beurteilung von Muedigkeit einfach!“ (Vollrath, 2018a) bietet hierfür eine präzise Analysemöglichkeit auf Basis verschiedener Müdigkeitsindikatoren (für Details siehe Abschnitt 2.3.5.3). Um Vergleiche zwischen der Alpha-Spindelbasierten Ermüdungsmessung und dem videobasierten Müdigkeitsrating ziehen zu können, wurden die zu ratenden Automatisierungsabschnitte am Versuchsplan festgemacht: Die Einteilung der automatisierten Fahrdauer (UV 1) in acht Faktorstufen bildete die Grundlage für die Definition von acht einminütigen Videosequenzen, hinsichtlich derer die Müdigkeit der Versuchsteilnehmer beurteilt werden sollte. Die Gesamtlänge der Versuchsfahrt – und damit des Videos – variierte zwischen den einzelnen Probanden aufgrund von Schwankungen in der Fahrgeschwindigkeit (vgl. Abschnitt 3.1.1). Daher war ebenfalls für das Videomaterial notwendig, das zeitlich synchron zum EEG aufgezeichnet wurde, den Mittelpunkt eines jeden Zeitbereichs zu finden. Um diesen Mittelpunkt herum wurde dann ein einminütiges Fenster definiert, in dem die Videosequenz beurteilt werden konnte. So war die zeitliche Kongruenz zwischen EEG-Alpha-Spindelrate und videobasiertem Müdigkeitsrating probandenindividuell in der Datenanalyse gewährleistet.

3.1.8 Immersionsgüte

Gegenwärtig werden verschiedene Forschungsvorhaben im Bereich des automatisierten Fahrens unter Zuhilfenahme eines Wizard-of-Oz-Fahrzeugs durchgeführt (vgl. Abschnitt 2.4.4). Um die Immersionsgüte des methodischen Verfahrens zwischen verschiedenen Fahrzeugkonstruktionen vergleichend einordnen zu können (vgl. Müller et al., 2019), empfiehlt es sich, verschiedene Kriterien heranzuziehen. Im Rahmen des BMWI-geförderten Forschungsprojekts „Ko-HAF“ fand ein „Wizard-of-Oz-Workshop“ statt, der auf eine Standardisierung der Methode abzielte. Die Projektpartner verständigten sich, einerseits die Glaubwürdigkeit der Täuschung im Anschluss an den Fahrversuch abzufragen (vgl. Abschnitt 3.1.4.4) und andererseits als objektives Maß die Steuerungsqualität des Wizard-Fahrers anhand der „Standard Deviation of Lateral Position“ (SDLP) über die Versuchsdauer zu erfassen.

3.1.9 Inferenzstatistische Analyse

Die Nullhypothese wurde mit einer α -Fehlerwahrscheinlichkeit von fünf Prozent geprüft. Auch wenn die Nullhypothese nicht verworfen werden konnte, ist (in den meisten Fällen) η_p^2 oder Cohen's d (Cohen, 1988) als relatives Effektstärkemaß ausgewiesen, um die Größe des erklärten Varianzanteils in Relation zur Gesamtvarianz (einschließlich der unerklärten Fehlervarianz) zu beschreiben. Grundsätzlich werden alle Hypothesen zweiseitig getestet, auch wenn zuvor eine gerichtete Hypothese aufgestellt wurde. Falls die α -Fehlerwahrscheinlichkeit von .05 bei zweiseitiger Testung einer gerichteten Hypothese nicht unterschritten wurde, bei einseitiger Testung jedoch schon, wird das Ergebnis des einseitigen Hypothesentests mit dem Zusatz „einseitig getestet“ berichtet. Abstufungen der Unterschreitung der α -Fehlerwahrscheinlichkeit werden in „ $p < .05$ “, „ $p < .01$ “ oder „ $p < .001$ “ angegeben. Wenn die α -Fehlerwahrscheinlichkeit von .05 nicht unterschritten wurde, wird der exakte p -Wert berichtet. Alle inferenzstatistischen Analysen wurden mit Hilfe der Software „SPSS Statistics“ (Version 21; IBM SPSS, 2012) durchgeführt. Diese Angaben gelten auch für die Analyseschritte des in Abschnitt 4 vorgestellten Experiments 2.

3.2 Ergebnisse

3.2.1 Deskriptive Daten

Auf Basis der in Abschnitt 3.1.4.3 erklärten Vorgehensweise wurde die erhobene Stichprobe von $N = 35$ Probanden auf Basis der gemessenen EEG-Alpha-Spindelrate klassifiziert, ob sie ermüdeten oder nicht ermüdeten. Die Ergebnisse dieser Clusterung werden in Abschnitt 3.2.3.1 vorgestellt. Tabelle 7 gibt vorab eine Übersicht der geclusterten Teilstichproben in Bezug auf die erlebte Funktionszuverlässigkeit, des Geschlechts, des Alters, der Fahrerfahrung und der Fahrleistung pro Jahr. Den Koffeinkonsum abhängig von der erlebten Funktionszuverlässigkeit und der Ermüdungsklassifikation zeigt Tabelle 8.

Tabelle 7: Übersicht über die Stichprobenzusammensetzung: Probandenanzahl, Mittelwert (M) und Standardfehler (SE) für unterschiedliche Variablen separiert nach EEG-basierter Ermüdungsklassifikation.

Variable		„ermüdet“	„nicht ermüdet“
Funktionszuverlässigkeit	hoch	9	9
	niedrig	9	8
Geschlecht	weiblich	6	10
	männlich	12	7
Alter in Jahren	M	36	35
	SE	3	3
Fahrerfahrung in Jahren	M	18	17
	SE	3	3
Fahrleistung in km / Jahr	M	12600	15200
	SE	3400	2800

Die Schlafdauer vor dem Fahrversuch in Abhängigkeit von der Ermüdungsklassifikation ist in Abbildung 27 dargestellt. Ein t -Test für unabhängige Stichproben zeigte entgegen der Hypothese keinen signifikanten Unterschied in der Schlafdauer vor dem Fahrversuch zwischen „ermüdet“ und „nicht ermüdet“ klassifizierten Probanden ($t(33) = -.50$, $p = .62$, $\eta_p^2 = .01$).

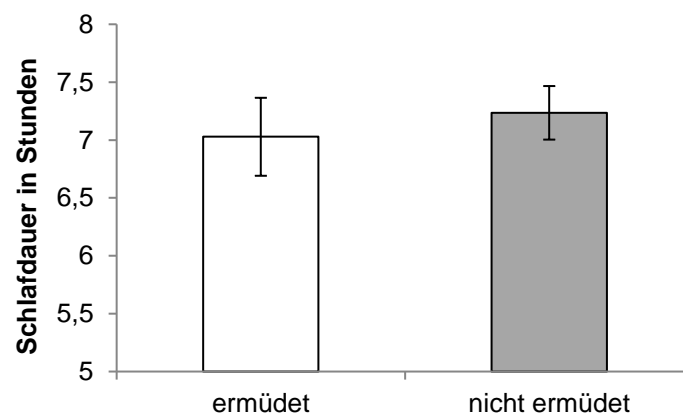


Abbildung 27: Mittlere Schlafdauer der Stichprobe vor dem Fahrversuch abhängig von der post hoc klassifizierten Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

Tabelle 8 listet die Anzahl der Probanden, die Koffein bzw. kein Koffein vor dem Versuch konsumiert haben, zugeordnet zur erlebten Funktionszuverlässigkeit und ihrer klassifizierten Ermüdung auf. Ein Chi-Quadrat-Unabhängigkeitstest zeigte, dass der Koffeinkonsum der Probanden unabhängig von der erlebten Funktionszuverlässigkeit war ($\chi^2(1) = .72$, $p = .40$, $\phi = .40$). Nur knapp wurde das Signifikanzniveau bei der Überprüfung der Unabhängigkeit zwischen Koffeinkonsum und klassifizierter Ermüdung entgegen der Hypothese verpasst ($\chi^2(1) = 2.33$, $p = .07$, $\phi = .13$, einseitig getestet).

Tabelle 8: Anzahl der Probanden, die Koffein oder kein Koffein vor dem Fahrversuch konsumiert hatten, abhängig von der erlebten Funktionszuverlässigkeit und der post hoc klassifizierten Ermüdung.

Variable	Koffein	kein Koffein
hohe Zuverlässigkeit	8	10
niedrige Zuverlässigkeit	10	7
„ermüdet“	7	11
„nicht ermüdet“	11	6

In der Vorbefragung (vgl. Abschnitt 3.1.1) wurde die Vorerfahrung der Probanden mit kontinuierlicher Fahrzeugautomatisierung (Wirkweise B, SAE Level 1) erhoben. Diese sind hinsichtlich Quer- und Längsregelung zu unterscheiden. Insgesamt 13 Probanden gaben an, regelmäßig einen Tempomaten oder einen Abstandsregeltempomaten (ACC) zu verwenden. Drei Probanden nutzten regelmäßig Spurhalte- oder Parklenkfunktionen (alle der Bedingung „niedrige Funktionszuverlässigkeit“ zufällig zugewiesen). Acht Längsführungsnutzer waren der hohen Zuverlässigkeit zugeordnet, hiervon alle vier ACC-Nutzer. Insgesamt sieben Längsführungsnutzer (drei verwenden hiervon ACC) und ein Querführungsnutzer wurden „ermüdet“ klassifiziert. Hiervon waren fünf Probanden der hohen Zuverlässigkeit zugeteilt, drei der niedrigen Zuverlässigkeit.

3.2.2 Subjektive Urteile

Ausschließlich ein Proband der Gesamtstichprobe äußerte auf die Testfrage „Auf welche Art und Weise wurde Ihrer Einschätzung nach das teilautomatisierte Fahren in diesem Versuch technisch umgesetzt?“, dass ggf. „ein zweiter Fahrer“ involviert gewesen sei. Alle anderen Versuchsteilnehmer gaben an, dass beispielsweise eine „Programmierung für die Fahrt auf der Teststrecke“ vorlag oder dass dem Fahrzeug „mittels Sensoren oder Kameras“ eine annähernd perfekte Automatisierung möglich war. Insofern kann davon ausgegangen werden, dass die Täuschung mittels Wizard-of-Oz-Methodik erfolgreich war.

3.2.2.1 Müdigkeitseinschätzung

Abbildung 28 zeigt die mittlere subjektive Müdigkeit der Probanden vor und nach der Versuchsfahrt in Abhängigkeit von der erlebten Funktionszuverlässigkeit. Außerdem ist in Abbildung 28 die mittlere subjektive Müdigkeit vor und nach der Versuchsfahrt abhängig von der klassifizierten Ermüdung auf Basis der EEG-Daten dargestellt. Ein t -Test für unabhängige Stichproben ergab zunächst, dass sich die subjektive Müdigkeit zu Versuchsbeginn nicht zwischen hoher und niedriger erlebter Funktionszuverlässigkeit unterschied ($t(30) = -.31, p = .76, \eta_p^2 < .01$).

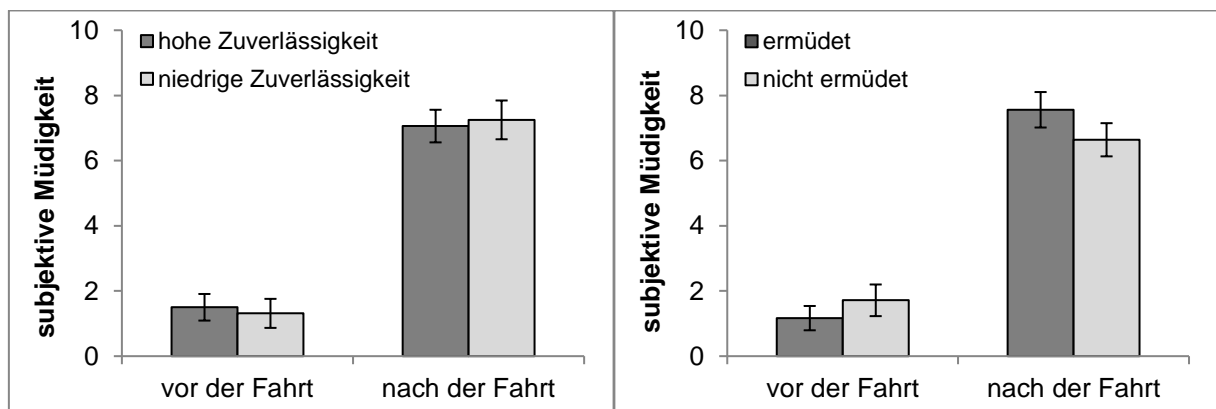


Abbildung 28: Mittlere subjektive Müdigkeit (SOMS) der Probanden abhängig vom Messzeitpunkt, links für eine hohe und niedrige erlebte Funktionszuverlässigkeit, rechts für die post hoc klassifizierte Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

Eine $2 \times 2 \times 2$ faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf einem Faktor zeigte entgegen der Hypothese keinen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 28) = .01, p = .97, \eta_p^2 < .01$) sowie des Zwischengruppenfaktors „Ermüdung“ ($F(1, 28) = .11, p = .74, \eta_p^2 < .01$). Hypothesenkonform konnte ein signifikanter Haupteffekt des messwiederholten Faktors „Zeitpunkt“ ($F(1, 28) = 161.05, p < .001$,

$\eta_p^2 = .85$) nachgewiesen werden. Eine Interaktion zwischen den Faktoren „Zeitpunkt“ und „Funktionszuverlässigkeit“ ($F(1, 28) = .21, p = .65, \eta_p^2 < .01$), „Zeitpunkt“ und „Ermüdung“ ($F(1, 28) = 2.68, p = .11, \eta_p^2 = .09$) sowie zwischen „Zeitpunkt“, „Funktionszuverlässigkeit“ und „Ermüdung“ ($F(1, 28) = .11, p = .74, \eta_p^2 < .01$) konnte nicht gefunden werden. Demnach war eine signifikant höhere subjektive Müdigkeit nach dem Versuch im Vergleich zum Versuchsbeginn – unabhängig von der erlebten Funktionszuverlässigkeit oder der objektiv klassifizierten Ermüdung – festzustellen. Tendenziell zeigte sich eine höhere subjektive Ermüdung nach der Fahrt (unabhängig von der Funktionszuverlässigkeit) für die mittels EEG als „ermüdet“ klassifizierten Probanden.

In einem weiteren Schritt wurde die subjektive Müdigkeit in Abhängigkeit vom Koffeinkonsum vor und nach der Versuchsfahrt bestimmt (Abbildung 29). Sowohl vor der Fahrt ($t(30) = -.44, p = .67, \eta_p^2 = .01$) als auch nach der Fahrt ($t(33) = -1.05, p = .30, \eta_p^2 = .03$) war kein signifikanter Unterschied der mittels SOMS erfassten Müdigkeit zwischen Koffeinkonsum und keinem Koffeinkonsum entgegen der Erwartung festzustellen.

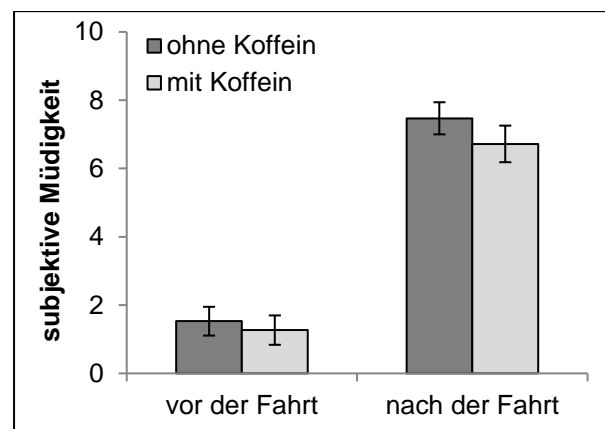


Abbildung 29: Mittlere subjektive Müdigkeit (SOMS) der Probanden abhängig vom Messzeitpunkt für Koffeinkonsum und keinen Koffeinkonsum. Die Fehlerbalken repräsentieren den Standardfehler.

3.2.2.2 Vertrauen in die Automatisierungsfunktion

In Abbildung 30 ist das mittlere Vertrauen der Probanden als Summenscore über alle Items der „Automation Trust Scale“ vor und nach der Versuchsfahrt in Abhängigkeit von der erlebten Funktionszuverlässigkeit dargestellt. Des Weiteren zeigt Abbildung 30 das mittlere Vertrauen der Probanden als Summenscore der „Automation Trust Scale“ über alle Items vor und nach der Versuchsfahrt abhängig von der klassifizierten Ermüdung auf Basis der EEG-Daten. Ein t -

Test für unabhängig Stichproben ergab zunächst, dass sich das Vertrauen in die Automatisierungsfunktion zu Versuchsbeginn nicht zwischen hoher und niedriger erlebter Funktionszuverlässigkeit ($t(33) = -.15, p = .88, \eta_p^2 < .01$), sowie zwischen den post hoc hinsichtlich ihrer Ermüdung klassifizierten Probanden ($t(33) = -1.07, p = .29, \eta_p^2 = .03$) unterschied.

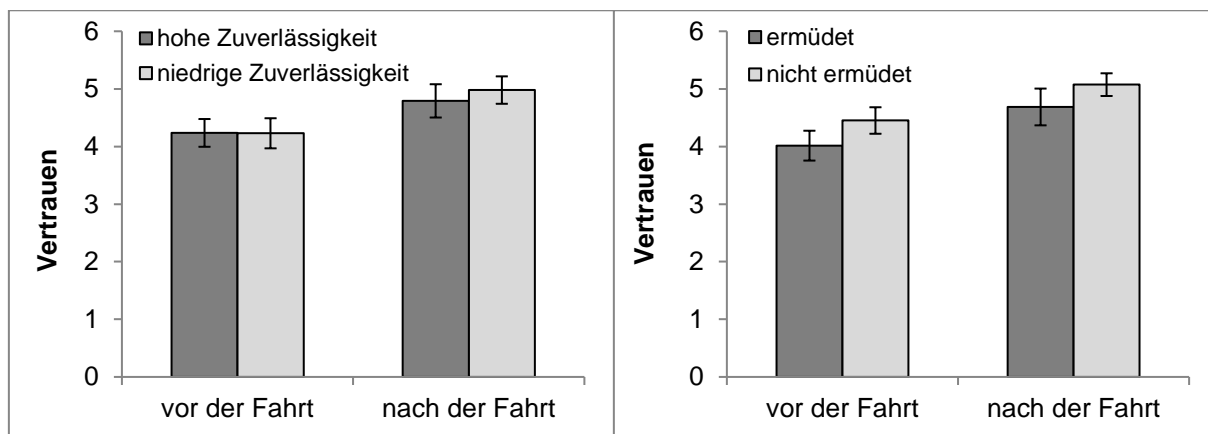


Abbildung 30: Mittlerer Summenscore der „Automation Trust Scale“ als Vertrauensmaß der Probanden abhängig vom Messzeitpunkt, links für eine hohe und niedrige erlebte Funktionszuverlässigkeit, rechts für die post hoc klassifizierte Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

Eine $2 \times 2 \times 2$ faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf einem Faktor zeigte entgegen der Hypothese keinen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 30) = .08, p = .78, \eta_p^2 < .01$) sowie des Zwischengruppenfaktors „Ermüdung“ ($F(1, 30) = 1.82, p = .19, \eta_p^2 = .06$) auf das Vertrauen in die Automatisierungsfunktion. Hypothesenkonform konnte ein signifikanter Haupteffekt des messwiederholten Faktors „Zeitpunkt“ ($F(1, 30) = 14.48, p < .001, \eta_p^2 = .33$) nachgewiesen werden. Eine Interaktion zwischen den Faktoren „Zeitpunkt“ und „Funktionszuverlässigkeit“ ($F(1, 30) = .32, p = .58, \eta_p^2 = .01$), „Zeitpunkt“ und „Ermüdung“ ($F(1, 30) = .01, p = .94, \eta_p^2 < .01$) sowie zwischen „Zeitpunkt“, „Funktionszuverlässigkeit“ und „Ermüdung“ ($F(1, 30) = 1.11, p = .30, \eta_p^2 = .04$) konnte nicht gefunden werden. Demnach war ein signifikant höheres Vertrauen in die Funktion nach dem Versuch im Vergleich zum Versuchsbeginn unabhängig von der erlebten Zuverlässigkeit oder der objektiv klassifizierten Ermüdung festzustellen.

3.2.2.3 Workload

Abbildung 31 zeigt den mittleren Gesamtworkload (über alle NASA-TLX-Skalen) in Abhängigkeit von der erlebten Funktionszuverlässigkeit und in Abhängigkeit von der klassifizierten Ermüdung auf Basis der EEG-Daten. Ein t -Test für unabhängige Stichproben zeigte entgegen der Erwartung keinen signifikanten Unterschied des eingeschätzten Workloads zwischen der

erlebten Funktionszuverlässigkeit ($t(33) = .07, p = .95, \eta_p^2 < .01$). Allerdings konnte ein t -Test für unabhängige Stichproben einen signifikant geringeren mittleren Gesamtworkload für die „ermüdet“ klassifizierten Probanden feststellen ($t(33) = -2.06, p < .05, \eta_p^2 = .11$).

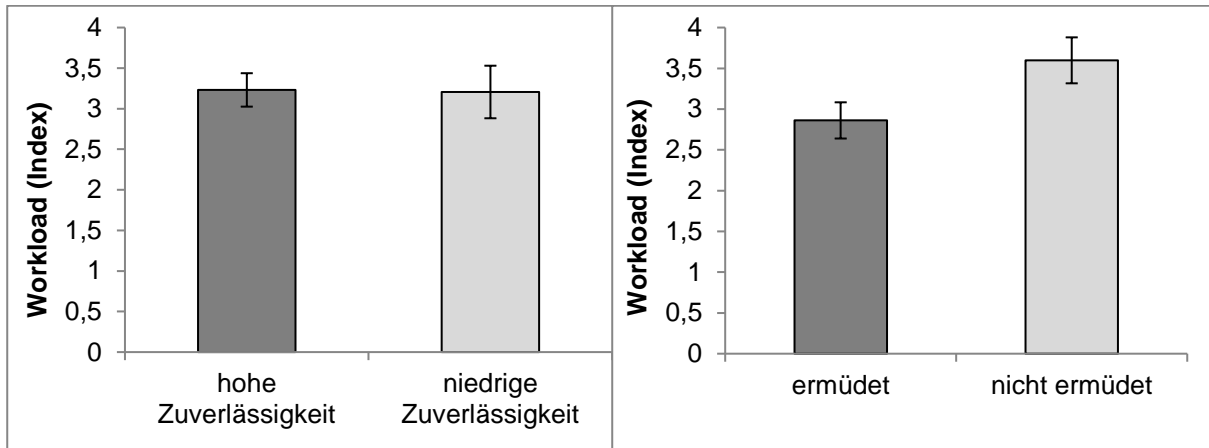


Abbildung 31: Mittlerer Gesamtworkload der Probanden über alle NASA-TLX-Items, links abhängig von der erlebten Funktionszuverlässigkeit, rechts abhängig von der post hoc klassifizierten Ermüdung. Ein Index von eins entspricht 10 % des maximal messbaren Workloads. Die Fehlerbalken repräsentieren den Standardfehler.

Um die Frage zu klären, auf welchen Subskalen des NASA-TLX dieser signifikante Unterschied auffindbar ist, wurde eine zweifaktorielle multivariate Varianzanalyse (MANOVA) für die sechs Subskalen berechnet. Die Analyse ergab, dass entsprechend der Erwartung ausschließlich eine signifikant niedrigere *kognitive* Beanspruchung der „ermüdeten“ Probanden vorlag ($F(1, 33) = 4.42, p < .05, \eta_p^2 = .12$).

3.2.2.4 Stresserleben

In Abbildung 32 ist der mittlere empfundene Stress der Probanden als Summenscore über alle Items vor und nach der Versuchsfahrt in Abhängigkeit von der erlebten Funktionszuverlässigkeit dargestellt. Ein t -Test für unabhängige Stichproben ergab, dass die Probanden in der Bedingung „hohe Zuverlässigkeit“ ein signifikant niedrigeres Stressniveau zu Versuchsbeginn aufwiesen als Probanden der Bedingung „niedrige Zuverlässigkeit“ ($t(33) = -2.96, p < .01, \eta_p^2 = .21$), obwohl die Gruppenzuweisung randomisiert erfolgte. Entgegen der Hypothese zeigte ein t -Test für abhängige Stichproben keinen signifikanten Unterschied zwischen dem empfundenen Stress vor und nach der Versuchsfahrt ($t(34) = 1.39, p = .18, \eta_p^2 = .06$). Wird ausschließlich der empfundene Stress nach der Versuchsfahrt betrachtet, ist entgegen der Hypothese kein signifikanter Unterschied zwischen der erlebten Funktionszuverlässigkeit festzu-

stellen ($t(28.12) = -1.61, p = .11, \eta_p^2 = .07$). A priori wurde davon ausgegangen, dass Versuchsteilnehmer mit hohem Stressempfinden höhere Neurotizismus-Ausprägungen besitzen. Die Korrelation zwischen dem Stressempfinden vor dem Fahrversuch und dem Persönlichkeitstrait lag erwartungsgemäß bei $r = .31$ ($p < .05$, einseitig getestet), zwischen dem Stressempfinden nach dem Fahrversuch und Neurotizismus bei $r = .15$ ($p = .38$).

Weiterhin konnte mit Hilfe eines t -Tests für unabhängige Stichproben zunächst festgestellt werden, dass sich der empfundene Stress zu Beginn des Versuchs nicht zwischen den post hoc als „ermüdet“ und als „nicht ermüdet“ klassifizierten Probanden unterschied ($t(33) = -1.22, p = .23, \eta_p^2 = .04$). Abbildung 32 zeigt ebenfalls den mittleren empfundenen Stress der Probanden als Summenscore über alle Items nach der Versuchsfahrt in Abhängigkeit von der klassifizierten Ermüdung auf Basis der EEG-Daten. Ein t -Test für unabhängige Stichproben zeigte entgegen der Hypothese keinen signifikanten Unterschied im Stressempfinden abhängig von der Ermüdungsstärke über die Versuchsfahrt ($t(33) = -1.47, p = .15, \eta_p^2 = .06$).

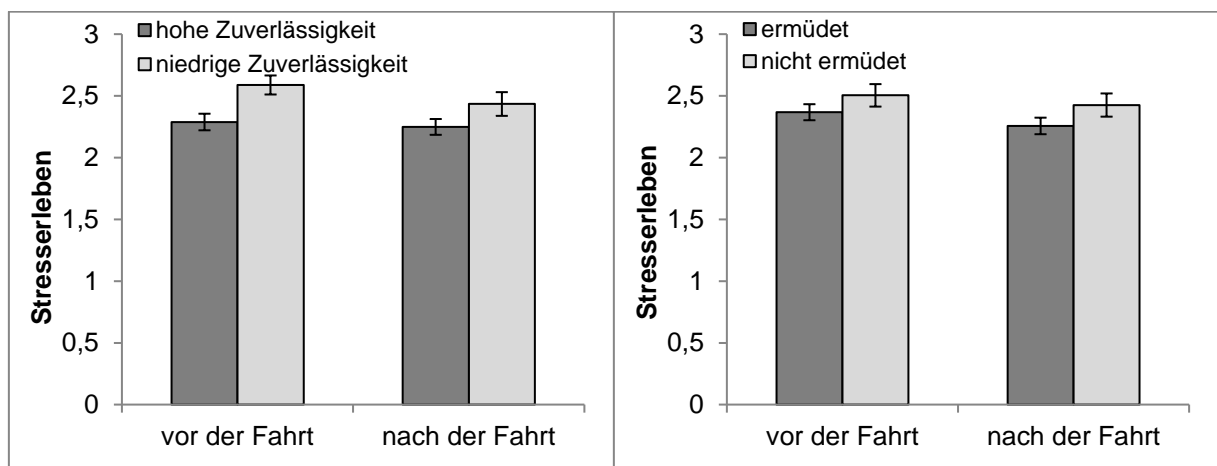


Abbildung 32: Mittleres Stresserleben der Probanden abhängig vom Messzeitpunkt, links für eine hohe und niedrige erlebte Funktionszuverlässigkeit, rechts für die post hoc klassifizierte Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

3.2.2.5 Nutzungsbereitschaft und Komforterleben

Abbildung 33 zeigt die mittlere Nutzungsbereitschaft der Probanden vor und nach der Versuchsfahrt in Abhängigkeit von der erlebten Funktionszuverlässigkeit. Zunächst wurde geprüft, ob bereits zu Beginn des Versuchs eine unterschiedliche Nutzungsbereitschaft zwischen den Bedingungen „hohe Zuverlässigkeit“ und „niedrige Zuverlässigkeit“ vorlag. Beide Bedingungen wiesen hierbei den gleichen Mittelwert auf ($t(32) = .00, p = 1.00, \eta_p^2 = .00$). Abbildung 33 zeigt ebenfalls die mittlere Nutzungsbereitschaft der Probanden vor und nach der Versuchsfahrt in Abhängigkeit von ihrer post hoc klassifizierten objektiven Ermüdung. Auch hier wurde geprüft, ob bereits zu Beginn des Versuchs eine unterschiedliche Nutzungsbereitschaft zwischen „ermüdeten“ und „nicht ermüdeten“ Probanden vorlag. Ein t -Test für unabhängige Stichproben ergab keinen signifikanten Unterschied ($t(32) = -1.10, p = .28, \eta_p^2 = .04$).

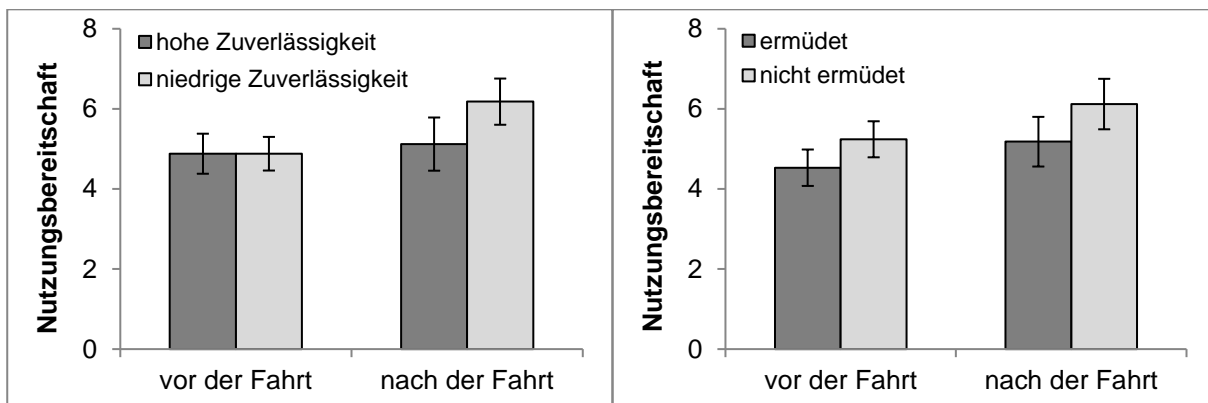


Abbildung 33: Nutzungsbereitschaft der Probanden abhängig vom Messzeitpunkt, links für eine hohe und niedrige erlebte Funktionszuverlässigkeit, rechts für die post hoc klassifizierte Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

Eine $2 \times 2 \times 2$ faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf einem Faktor zeigte weder einen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 32) = .56, p = .46, \eta_p^2 = .02$) noch einen signifikanten Haupteffekt des Zwischengruppenfaktors „Ermüdung“ ($F(1, 32) = 1.40, p = .25, \eta_p^2 = .04$) auf die Bereitschaft, die erlebte Automatisierungsfunktion im Alltag zu nutzen. Jedoch war ein signifikanter Haupteffekt des messwiederholten Faktors „Zeitpunkt“ ($F(1, 30) = 5.79, p < .05, \eta_p^2 = .16$) zu ermitteln. Eine Interaktion zwischen den Faktoren „Zeitpunkt“ und „Funktionszuverlässigkeit“ ($F(1, 30) = 2.75, p = .11, \eta_p^2 = .08$), „Zeitpunkt“ und „Ermüdung“ ($F(1, 30) = .21, p = .65, \eta_p^2 < .01$) sowie zwischen „Zeitpunkt“, „Funktionszuverlässigkeit“ und „Ermüdung“ ($F(1, 30) = .75, p = .39, \eta_p^2 = .02$) konnte nicht gefunden werden. Dementsprechend stieg die Nutzungsbereitschaft der Automatisierungsfunktion mit dem Erleben der Funktionsweise an, unabhängig von der erlebten Funktionszuverlässigkeit oder der aufgetretenen Müdigkeit.

In Abbildung 34 ist der mittlere erlebte Komfort abhängig von der erlebten Funktionszuverlässigkeit sowie der post hoc klassifizierten objektiven Ermüdung dargestellt. Ein t -Test für unabhängige Stichproben zeigte keinen signifikanten Unterschied des eingeschätzten Komforts zwischen hoher und niedriger Zuverlässigkeit ($t(33) = -.76, p = .45, \eta_p^2 = .02$) und zwischen beiden Ermüdungsklassen ($t(33) = -.63, p = .53, \eta_p^2 = .01$).

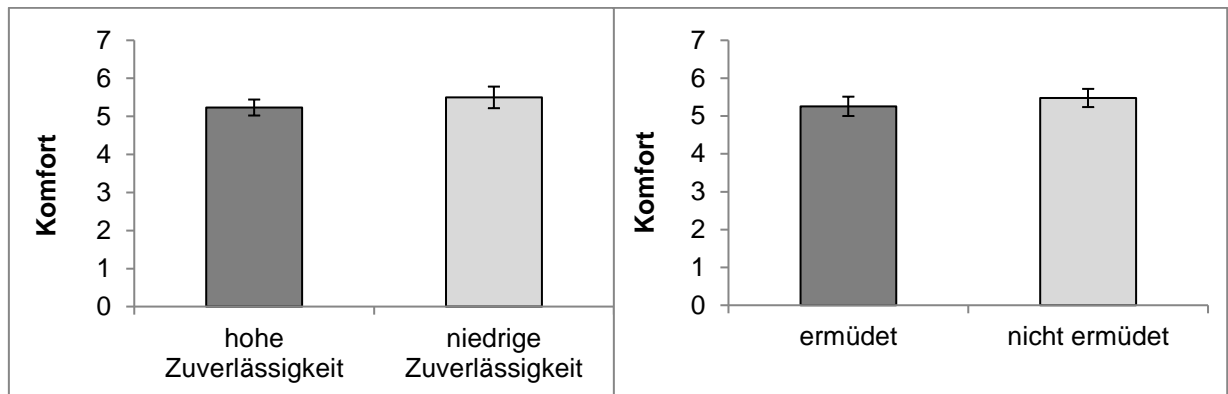


Abbildung 34: Mittlerer empfundenen Komfort der Probanden, links abhängig von der erlebten Funktionszuverlässigkeit, rechts abhängig von der post hoc klassifizierten Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

3.2.2.6 Kinetose

Insgesamt vier aller 35 Probanden wiesen entsprechend der von Golding (2006) definierten Norm (Mittelwert \pm eine Standardabweichung) eine erhöhte Anfälligkeit zur Entwicklung von Kinetose in der Vorbefragung auf. Im Mittel lag der MSSQ-Score bei 7.47 Punkten ($SE = 1.56, SD = 9.20$), wobei die Streuung der Werte zwischen null und 32.14 Punkten lag. Zwischen der in der Vorbefragung prädizierten Neigung zur Kinetose und dem in der Nachbefragung berichteten Befinden der Probanden konnte eine signifikante Korrelation (nach Pearson) gefunden werden ($r = .46, p < .01$).

Abbildung 35 zeigt einerseits den mittleren MSSQ-Score der Probanden in Abhängigkeit von der Ermüdungsklassifikation, und andererseits das mittlere subjektive Befinden im Anschluss an den Fahrversuch abhängig von der erlebten Funktionszuverlässigkeit und der klassifizierten Ermüdung. Ein t -Test für unabhängige Stichproben zeigte entgegen der Hypothese keinen signifikanten Unterschied in der Anfälligkeit zur Kinetose (MSSQ) und der post hoc klassifizierten Ermüdung ($t(33) = -1.21, p = .23, \eta_p^2 = .04$). Eine 2×2 faktorielle Varianzanalyse (ANOVA) konnte auch keinen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 31) < .01, p = .95, \eta_p^2 < .01$) sowie des Zwischengruppenfaktors „Ermüdung“

($F(1, 31) = 1.95, p = .17, \eta_p^2 = .06$) ermitteln. Eine Interaktion zwischen dem Faktor „Funktionszuverlässigkeit“ und „Ermüdung“ konnte ebenfalls nicht gefunden werden ($F(1, 31) = .45, p = .51, \eta_p^2 = .01$).

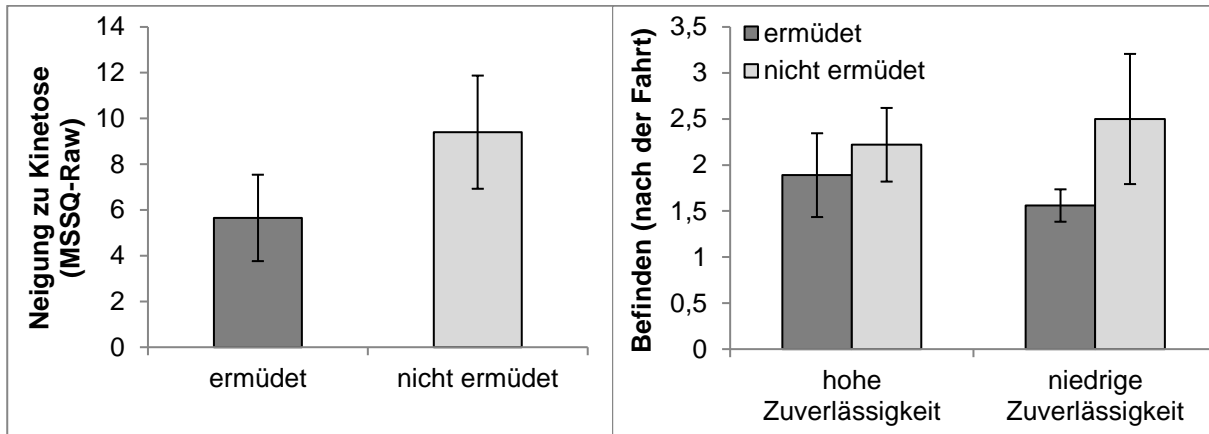


Abbildung 35: Links: Mittlere Neigung der Probanden zur Ausbildung von Kinetose abhängig von der post hoc klassifizierten Ermüdung. Rechts: Mittleres subjektives Befinden der Probanden nach der Versuchsfahrt abhängig von der erlebten Funktionszuverlässigkeit für post hoc klassifizierte Ermüdung und keine Ermüdung. Die Fehlerbalken repräsentieren den Standardfehler.

3.2.3 EEG-Alpha-Spindelrate

3.2.3.1 Zeitreihen

Für eine Gesamtübersicht der gemessenen Alpha-Spindeln aller Probanden – aufgeteilt auf die Bedingungen „hohe Funktionszuverlässigkeit“ und „niedrige Funktionszuverlässigkeit“ – ist die Rate für jede Minute automatisierter Fahrt (Level 2) in Abbildung 36 aufgetragen. Die Nummerierung der Probanden beginnt bei neun, da die ersten acht Probanden an verschiedenen Vorversuchen teilnahmen.

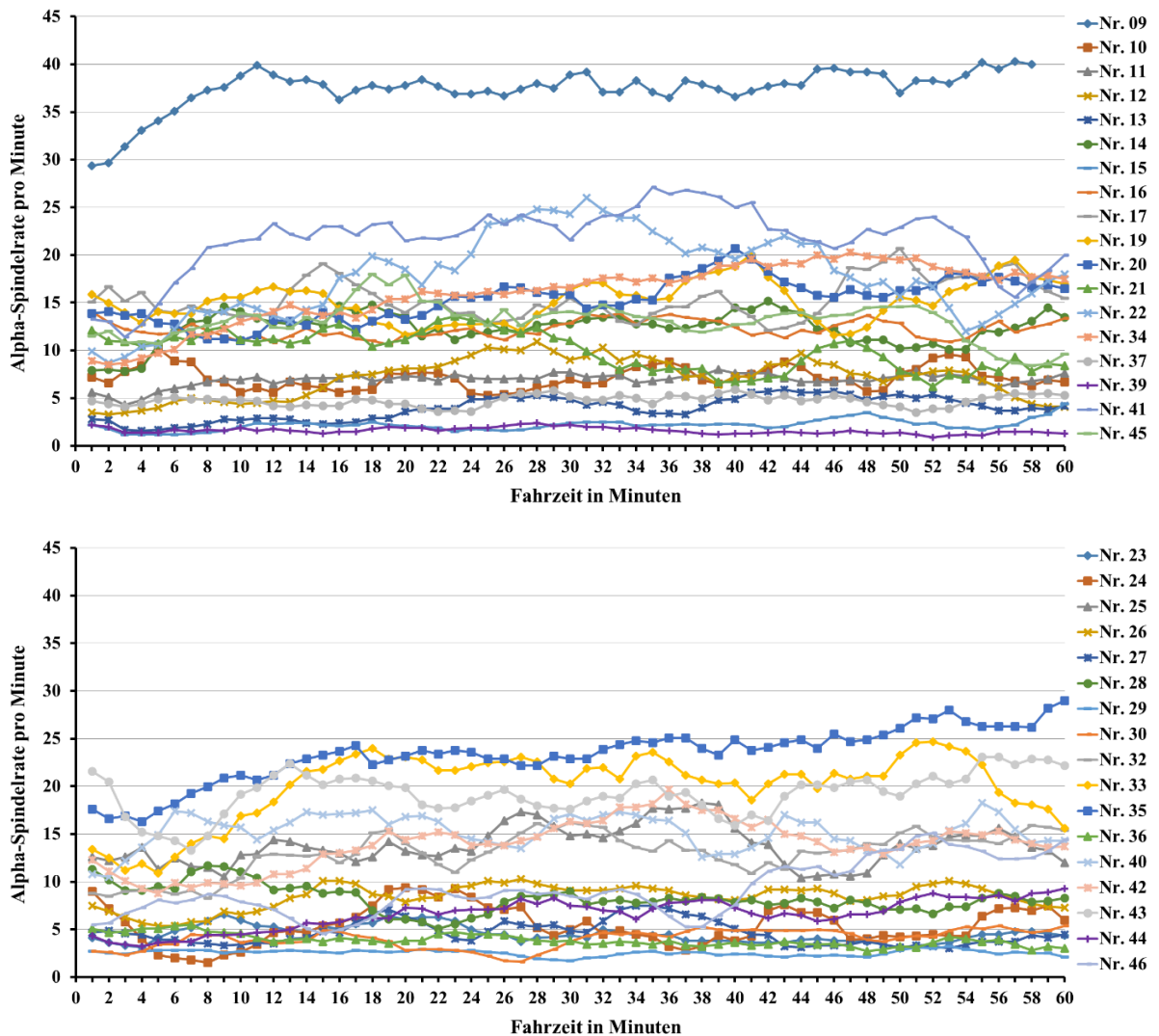


Abbildung 36: EEG-Alpha-Spindelrate pro Minute abhängig von der Fahrzeit. Jede Zeitreihe ist einem Probanden zugeordnet. Oben: hohe Funktionszuverlässigkeit. Unten: niedrige Funktionszuverlässigkeit.

Unabhängig von der Versuchsbedingung ist bei Betrachtung der Abbildung 36 die in Schmidt (2010) beschriebene Individualität der Alpha-Spindelrate wieder zu finden: Manche Probanden weisen in den Zeitreihen sehr niedrige Spindelraten (unter zehn Spindeln pro Minute) auf, andere gehen über 20 Spindeln pro Minute hinaus. In der Bedingung „hohe Funktionszuverlässigkeit“ fällt Proband Nr. 09 mit besonders hoher Spindelrate auf, dessen Zeitreihe sich zwischen 30 und 40 Spindeln pro Minute bewegt. Nach probandenindividueller Referenzierung (Abbildung 37; vgl. Abschnitt 3.1.4.3) ist auf der Ordinate folglich nicht mehr die absolute Alpha-Spindelrate eines Probanden, sondern die an Minute fünf referenzierte (relative) Rate aufgetragen, beginnend bei Minute fünf und endend bei Minute 60 (d.h. es gibt 55 Messzeitpunkte). So lässt sich anhand der Zeitreihen ablesen, um wie viele Alpha-Spindeln ein Anstieg oder eine Reduktion in Vergleich zu Minute fünf zu verzeichnen ist.

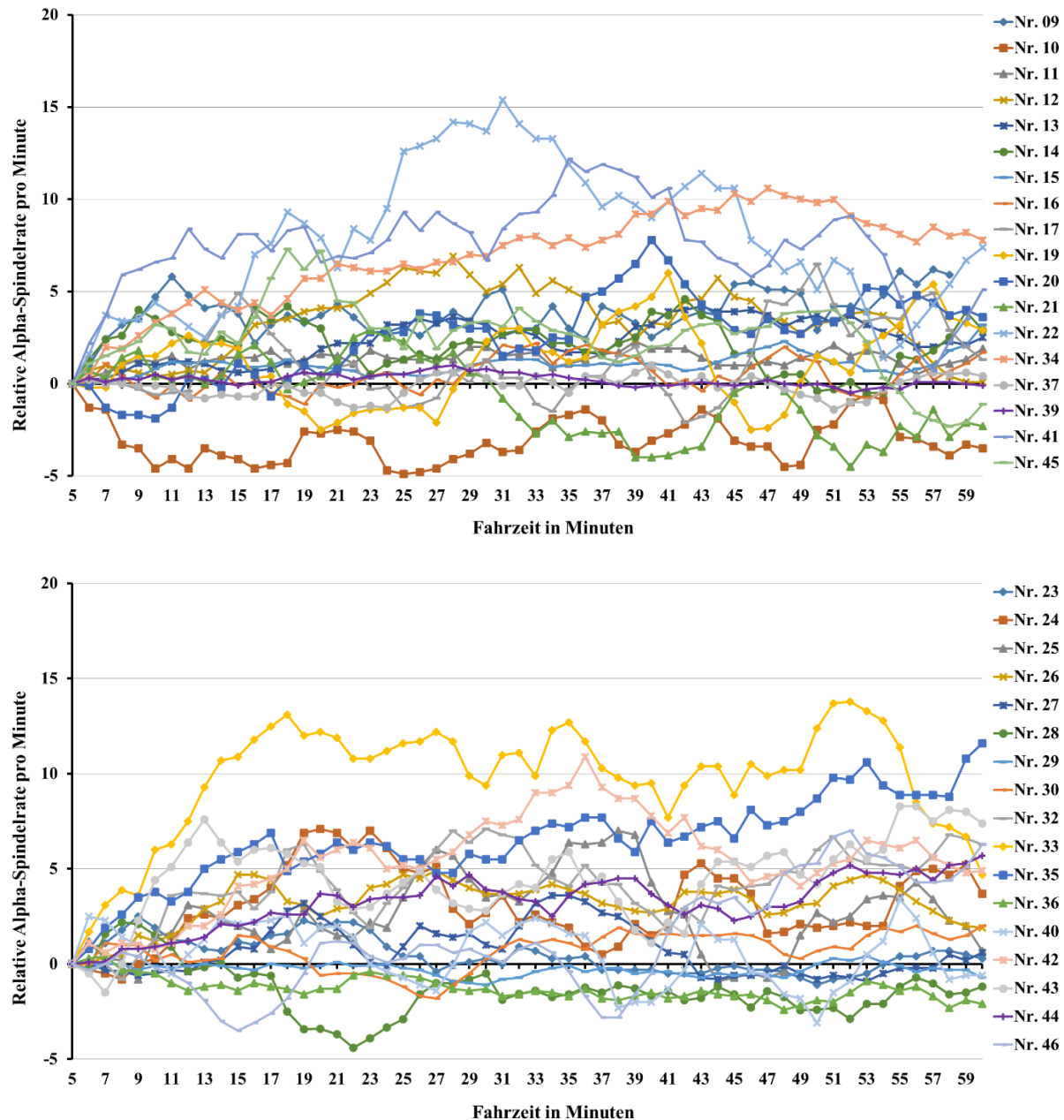


Abbildung 37: Relative EEG-Alpha-Spindelrate pro Minute abhängig von der Fahrzeit (referenziert an Minute 5). Jede Zeitreihe ist einem Probanden zugeordnet. Oben: hohe Funktionszuverlässigkeit. Unten: niedrige Funktionszuverlässigkeit.

Zunächst fällt beim Betrachten der Abbildung 37 auf, dass jeder Proband weiterhin einen höchst individuellen Verlauf der Alpha-Spindelrate aufweist. Nun lässt sich jedoch die relative Zu- oder Abnahme der Spindelrate zwischen den Probanden vergleichen: Einerseits steigt bei einigen Probanden die Alpha-Spindelrate über die Versuchsdauer an – teils mit rapidem zwischenzeitlichen Absinken. Andererseits scheint bei manchen Probanden über einen Großteil der Level 2-Fahrt eine Alpha-Spindelrate unterhalb des Ausgangswerts null vorzuliegen. Durchgängig treten Schwankungen der Rate auf, wodurch kein kongruentes Muster über alle Probanden

beim Ansteigen oder Abfallen der Rate zu erkennen ist. Insgesamt weisen einige Probanden nach 60 Minuten Level 2-Fahrt eine Alpha-Spindelrate über null, andere wiederum eine Alpha-Spindelrate unter null auf. In der Konsequenz existiert große interindividuelle Varianz zu jeden Messzeitpunkt – häufig beträgt sie zwischen 10 und 15 Spindeln pro Minute. Wie kann mit dieser hohen Inkonsistenz der Alpha-Spindelrate im zeitlichen Verlauf zwischen den Probanden umgegangen werden?

Für die nachfolgend erläuterte Vorgehensweise wurde entsprechend Abschnitt 2.3.5.4 davon ausgegangen, dass ein Anstieg der Alpha-Spindelrate mit einem Anstieg der Ermüdung eines Probanden einhergeht. Folglich wäre es nicht zielführend, weitere Berechnungen über die Gesamtstichprobe von $N = 35$ durchzuführen, weil einige Probanden scheinbar schneller und stärker ermüdeten als andere. Wieder andere scheinen annähernd keine Anzeichen von Ermüdung aufzuweisen, sondern sind im Verlauf der Versuchsfahrt wacher als zum Startzeitpunkt (Spindelrate < 0). Etwaige Effekte, die sich auf die experimentelle Manipulation rückführen lassen sollten (die wahre Varianzquelle also), wären durch ein erhebliches Maß an individueller Fehlervarianz unsichtbar. Vor diesem Hintergrund empfiehlt es sich, vermeintlich „ermüdete“ Probanden von „nicht ermüdeten“ Probanden zu trennen. Das Verfahren, welches hierfür zur Anwendung kam, ist in Abschnitt 3.1.4.3 erläutert. Dementsprechend zeigt Abbildung 38 zunächst die relative Alpha-Spindelrate zum jeweiligen Messzeitpunkt (Zeitreihe von Minute fünf bis Minute 60) für die „nicht ermüdet“ klassifizierten Probanden beider erlebter Funktionszuverlässigkeiten, wohingegen Abbildung 39 die relative Alpha-Spindelrate als Zeitreihe für die „ermüdet“ klassifizierten Probanden von beiden Zuverlässigkeitsbedingungen darstellt.

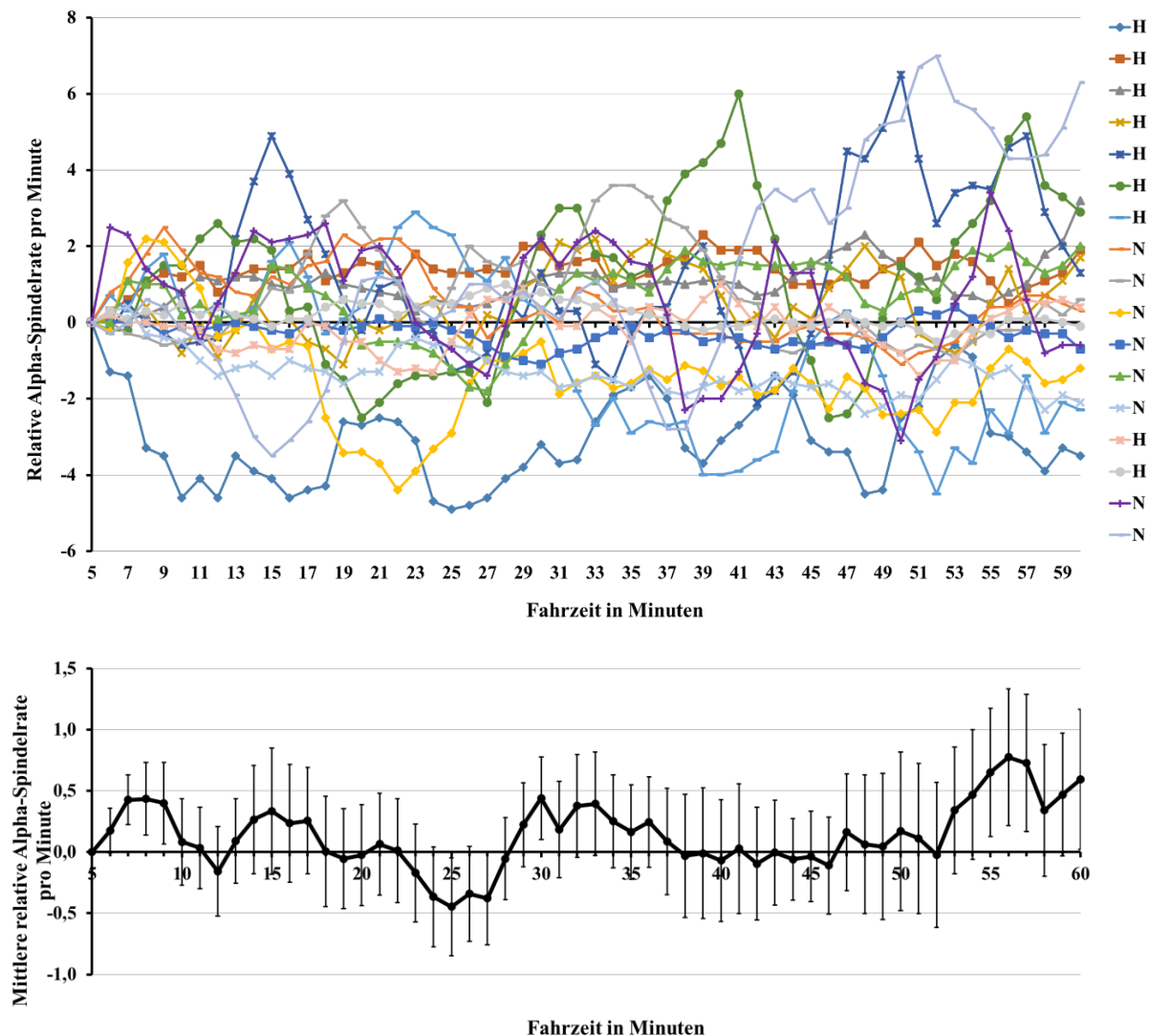


Abbildung 38: Oben: Relative EEG-Alpha-Spindelrate pro Minute für „nicht ermüdet“ klassifizierte Probanden abhängig von der Fahrzeit. Die Legende zeigt die erlebte Funktionszuverlässigkeit (H = hoch, N = niedrig). Unten: Mittlere relative EEG-Alpha-Spindelrate über alle „nicht ermüdet“ klassifizierten Probanden abhängig von der Fahrzeit. Die Fehlerbalken repräsentieren den Standardfehler.

Anhand Abbildung 38 ist zu erkennen, dass die mittlere Alpha-Spindelrate der „nicht ermüdeten“ Probanden lediglich im Bereich von einer Spindel pro Minute schwankt, mit sehr großer interindividueller Varianz. Eine Alpha-Spindelrate kleiner null bedeutet hier, dass die Probanden im Mittel wacher waren als zu Beginn der Level 2-Fahrt (d.h. zum Referenzzeitpunkt „Minute 5“). Im Gegensatz zu den „nicht ermüdeten“ Probanden ist für die „ermüdet“ klassifizierte Teilstichprobe ($n = 18$) eine im Mittel ansteigende Alpha-Spindelrate bis etwa Minute 20 erkennbar. Die Zunahme beträgt zu diesem Zeitpunkt mehr als durchschnittlich fünf Spindeln pro Minute, vgl. Abbildung 39.

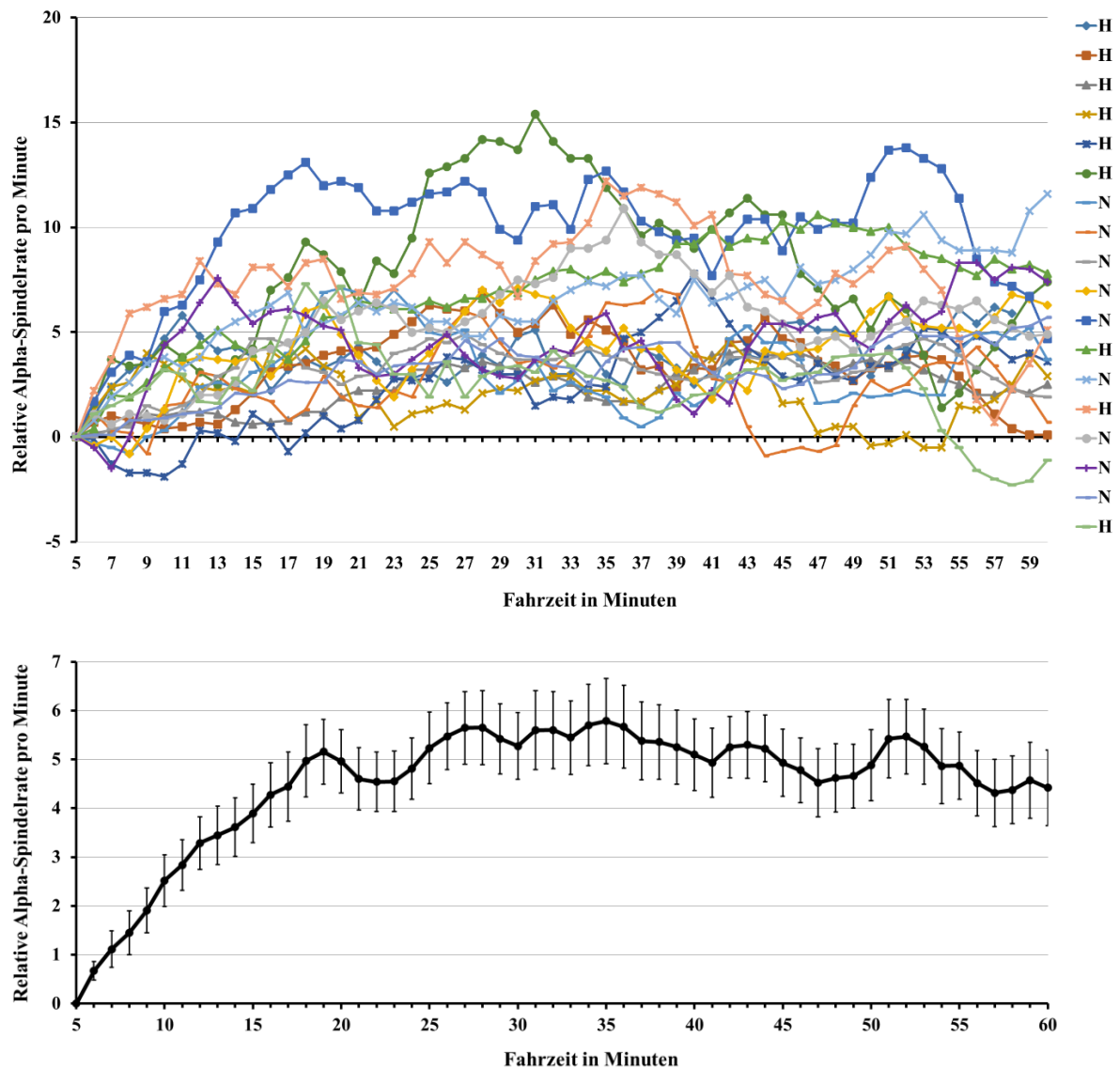


Abbildung 39: Oben: Relative EEG-Alpha-Spindelrate pro Minute für „ermüdet“ klassifizierte Probanden abhängig von der Fahrzeit. Die Legende zeigt die erlebte Funktionszuverlässigkeit (H = hoch, N = niedrig). Unten: Mittlere relative EEG-Alpha-Spindelrate über alle „ermüdet“ klassifizierten Probanden abhängig von der Fahrzeit. Die Fehlerbalken repräsentieren den Standardfehler.

Die Daten sollten entsprechend der Fragestellungen (vgl. Abschnitt 2.5.1) näher untersucht werden, um herauszufinden, wie groß der Einfluss der Fahrdauer mit einer Level 2-Funktion auf die Ermüdung ist. Hierfür wurde der Zeitbereich von Minute fünf bis Minute 20 (d.h. über 16 Datenpunkte) gewählt, in dem die Ermüdung monoton steigt – es lässt sich ein quasi-linearer Zusammenhang vermuten. Wird ausschließlich ein lineares Regressionsmodell mit einem Prädiktor – der Fahrdauer – berechnet, so zeichnet sich eine hohe mittlere Anpassungsgüte nach den Richtlinien von Cohen (1988) bzw. Varianzaufklärung von $R^2 = .66$ ($SE = .06$) ab. Dieser einzige Prädiktor ist bei 17 der 18 Probanden in der Lage, das Kriterium (Zunahme der Alpha-

Spindelrate innerhalb der ersten 20 Minuten Fahrt) signifikant vorherzusagen. Tabelle 9 zeigt die Regressionskoeffizienten (R), den aufgeklärten Varianzanteil (R^2), die Teststatistik t und die zugehörigen p -Werte für jeden Probanden.

Tabelle 9: Ergebnisse einer linearen Regressionsanalyse über die ersten 20 Minuten Level 2-Fahrt für alle „ermüdet“ klassifizierten Probanden.

Proband	R / Beta	R^2	t	Effekt
09	.40	.16	1.61	$p = .13$
12	.89	.79	7.34	$p < .001$
13	.69	.48	3.59	$p < .01$
14	.59	.35	2.76	$p < .05$
20	.57	.32	2.58	$p < .05$
22	.88	.77	6.75	$p < .001$
24	.95	.91	11.97	$p < .001$
25	.56	.31	2.50	$p < .05$
26	.83	.69	5.57	$p < .001$
32	.91	.83	8.25	$p < .001$
33	.97	.94	15.15	$p < .001$
34	.89	.80	7.41	$p < .001$
35	.90	.81	7.66	$p < .001$
41	.77	.59	4.52	$p < .001$
42	.95	.90	11.32	$p < .001$
43	.78	.60	4.59	$p < .001$
44	.98	.96	17.40	$p < .001$
45	.87	.75	6.53	$p < .001$

Es wurde vermutet, dass ein nahezu perfektes Steuerungsverhalten des Versuchsleiters die Ermüdung der Probanden begünstigen könnte. Die Standardabweichung der Spurhaltung des Versuchsfahrzeugs (SDLP) in Bezug auf die linke Fahrstreifenmarkierung lag für jede Minute automatisierter Fahrt vor und lässt indirekt Rückschlüsse auf die Steuerungsgüte des Versuchsleiters zu (vgl. Abschnitt 2.3.5.2). Um einen Einfluss des lateralen Regelungsverhaltens vom

Wizard-Fahrer auf die Fahrermüdigkeit der Probanden zu überprüfen, wurde die SDLP für jeden Versuchsteilnehmer aus der Regressionsanalyse auspartialisiert. Hierfür ist in Tabelle 10 die Korrelation (nach Pearson) zwischen der Alpha-Spindelrate und der Fahrminute (analog zu Tabelle 9) und der SDLP für jeden „ermüdet“ klassifizierten Probanden dargestellt. Des Weiteren kann die Korrelation zwischen der Alpha-Spindelrate und der Fahrminute bei Auspartialisierung der SDLP abgelesen werden. Da der Effekt der Auspartialialisierung sehr klein ist, wurden an dieser Stelle drei Dezimalstellen angegeben. Nur bei einem der 18 Probanden konnte eine signifikante Korrelation (nach Pearson) zwischen der Alpha-Spindelrate und der SDLP gefunden werden ($r = -.60$, $p < .01$, einseitig getestet). Über alle Probanden hinweg ist dieser Effekt zufallskritisch nicht nachweisbar ($M_r = -.10$, $SE = .05$). Deskriptiv besteht jedoch ein hypothesenkonformer Trend, wonach eine geringere Steuerungsgüte mit einer Verringerung des Ermüdungsanstiegs einhergehen kann.

Tabelle 10: Korrelation zwischen Alpha-Spindelrate (SR) und SDLP, SR und Fahrminute sowie SR und Fahrminute unter Auspartialisierung der SDLP für alle „ermüdet“ klassifizierten Probanden.

Proband	$R_{\text{SR-SDLP}}$	$R_{\text{SR-Minute}}$	$R_{\text{SR-Minute(partial)}}$
09	-.142	.395	.399
12	.063	.891	.899
13	.080	.693	.696
14	.088	.593	.591
20	-.597	.567	.563
22	.081	.875	.875
24	.036	.954	.963
25	-.285	.556	.656
26	-.238	.830	.824
32	.135	.911	.910
33	.392	.971	.967
34	-.195	.893	.890
35	-.238	.898	.898
41	-.135	.770	.765
42	-.029	.950	.950

43	-.371	.775	.760
44	-.198	.978	.981
45	-.244	.867	.866

In einem weiteren Schritt wurde eine Aufteilung der „ermüdeten“ Probanden in die hohe und niedrige erlebte Funktionszuverlässigkeit vorgenommen und die mittlere relative Alpha-Spindelrate separat pro Bedingung aufgetragen, siehe Abbildung 40. Eine grobe Skizzierung der Zeitpunkte, an denen Probanden ein Automatisierungsfehler präsentiert wurde, soll helfen, etwaige Aktivierungseffekte von Automatisierungsfehlern (detailliert in Abschnitt 3.2.3.3 betrachtet) zunächst auf deskriptiver Ebene ausmachen zu können. Die zeitliche Unsicherheit dieser Skizzierung ist insofern groß, als dass für das Versuchsdesign zwar ein Abstand von acht Minuten zwischen jedem Automatisierungsfehler veranschlagt war, jedoch die realen Abstände (aufgrund teilweise höherer Fahrgeschwindigkeit und damit kürzerer Rundenzeit) kleiner waren. Diese Unschärfe kumuliert mit zunehmender Fahrdauer. In der Bedingung „hohe Funktionszuverlässigkeit“ sinkt die Alpha-Spindelrate im Mittel ab, nachdem der einzige Automatisierungsfehler präsentiert wurde. Die Bedingung „niedrige Funktionszuverlässigkeit“ zeigt durchschnittlich größere Schwankungen in der Alpha-Spindelrate, wobei ein Absinken der Rate häufig nach einem Automatisierungsfehler zu beobachten ist. Verglichen mit der hohen Zuverlässigkeit steigt bei niedriger Zuverlässigkeit die Spindelrate tendenziell an, nachdem im Bereich von Minute 50 ein Automatisierungsfehler erlebt wurde.

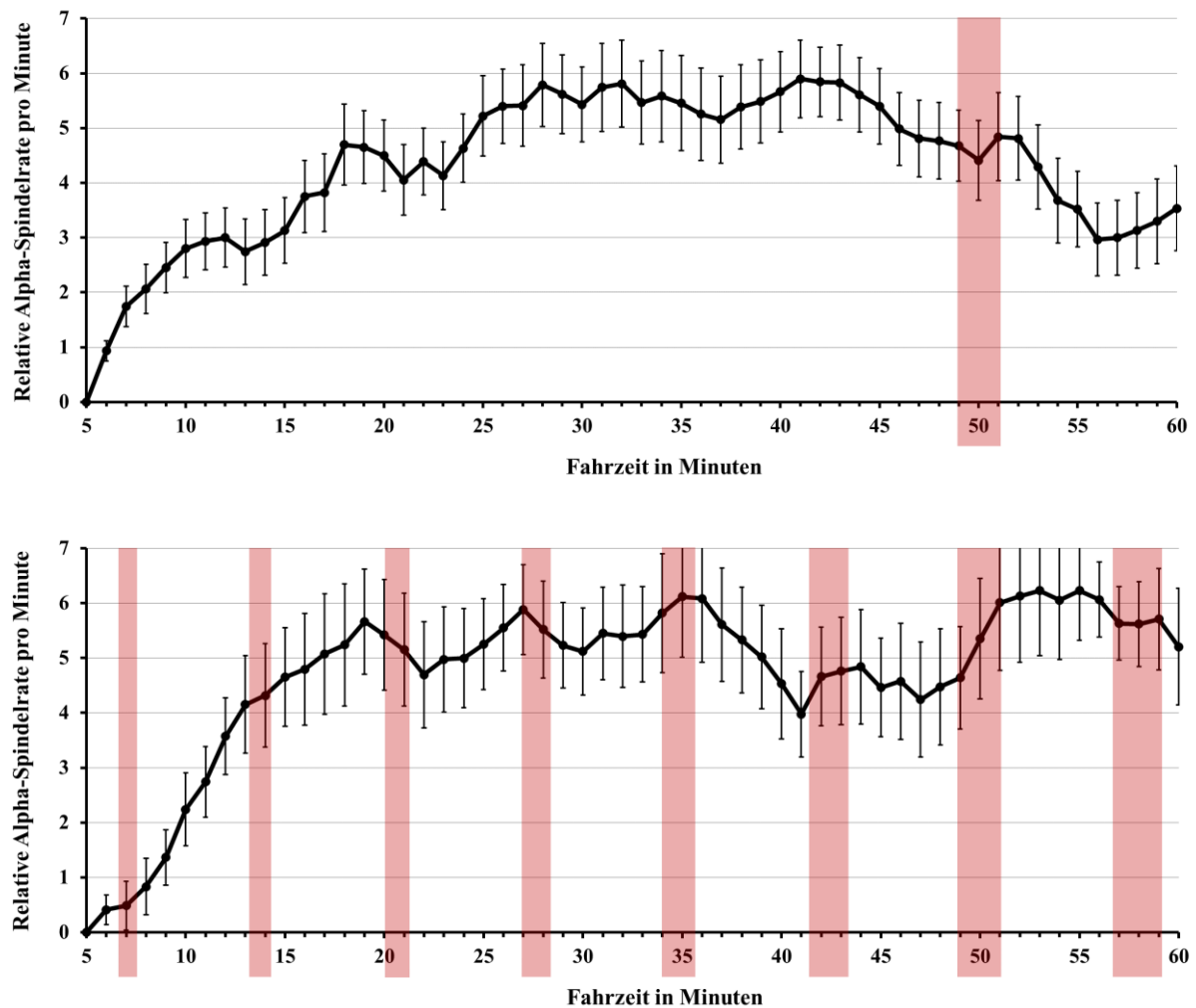


Abbildung 40: Oben: Relative EEG-Alpha-Spindelrate pro Minute für „ermüdet“ klassifizierte Probanden abhängig von der Fahrzeit bei hoher Funktionszuverlässigkeit. Der einzige Automatisierungsfehler (mit zeitlicher Unsicherheit) um Minute 50 ist grob skizziert. Unten: Mittlere relative EEG-Alpha-Spindelrate über alle „ermüdet“ klassifizierte Probanden abhängig von der Fahrzeit bei niedriger Funktionszuverlässigkeit. Alle acht Automatisierungsfehler (mit kumulierender zeitlicher Unsicherheit) sind grob skizziert. Die Fehlerbalken repräsentieren den Standardfehler.

3.2.3.2 Analyse über Aggregationsbereiche

Abbildung 41 stellt die mittlere Differenz der Alpha-Spindelrate für alle acht Aggregationsbereiche der „ermüdeten“ und „nicht ermüdeten“ Probanden für die hohe und niedrige erlebte Funktionszuverlässigkeit dar. Von Aggregationsbereich eins bis einschließlich sechs ist ein Ermüdungsanstieg anzunehmen, wobei die Messwerte in Aggregationsbereich sieben und acht auf eine Müdigkeitsreduktion hindeuten. Weiterhin ist über allen aggregierten Messwerten eine große Varianz (teils beträgt der Standardfehler mehr als eine Alpha-Spindel pro Minute) zu

beobachten, die auch hier mit steigender Ermüdung zunimmt. Eine 2×8 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf dem zweiten Faktor zeigte entgegen der Hypothese keinen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 16) = .18, p = .68, \eta_p^2 = .01$). Folglich unterscheiden sich beide Zuverlässigkeitsbedingungen nicht signifikant in ihrem Ermüdungsverhalten. Betrachtet man jedoch die einzelnen acht aggregierten Messzeitpunkte der hohen erlebten Zuverlässigkeit, so ist ein signifikanter Haupteffekt entsprechend der Hypothese festzustellen ($F(3.51, 56.09) = 21.36, p < .001, \eta_p^2 = .57$). Die Freiheitsgrade wurden nach Greenhouse-Geisser korrigiert mit $\epsilon = .50$.

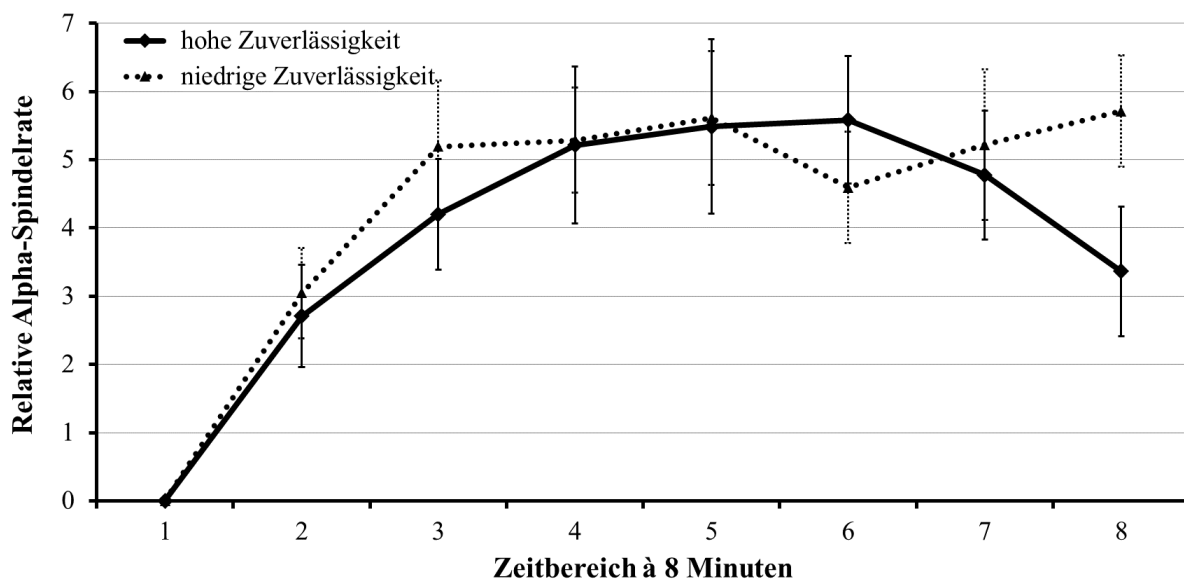


Abbildung 41: Relative EEG-Alpha-Spindelrate für „ermüdet“ klassifizierte Probanden der hohen und niedrigen erlebten Funktionszuverlässigkeit abhängig von der Fahrdauer (zeitlich aggregiert). Die Fehlerbalken repräsentieren den Standardfehler.

Zunächst wurde der deskriptiv zu erkennende Anstieg der Alpha-Spindelrate von Aggregationsbereich eins bis vier der hohen Zuverlässigkeitsbedingung analysiert. Vergleiche zwischen den Aggregationsbereichen vier, fünf und sechs wurden aufgrund der marginalen Messwerterhöhung zwischen einander – bei gleichzeitig höchster interindividueller Varianz innerhalb der aggregierten Messwerte – bewusst ausgespart. Demnach weisen post hoc Vergleiche darauf hin, dass die Alpha-Spindelrate im Aggregationsbereich zwei signifikant höher ist als im Aggregationsbereich eins ($p < .01$, einseitig getestet), sowie die Alpha-Spindelrate im Aggregationsbereich drei ebenfalls höher ist als im Aggregationsbereich zwei ($p < .01$, einseitig getestet). Im Aggregationsbereich vier konnte verglichen mit Aggregationsbereich drei kein signifikanter Anstieg der Alpha-Spindelrate mehr gemessen werden ($p = .10$, einseitig getestet). Dennoch interessierte, ob sich der deskriptiv höchste aggregierte Wert im Aggregationsbereich vier vom

Bezugswert null unterschied. Dieser post hoc Vergleich zwischen Aggregationsbereich vier und Aggregationsbereich eins zeigte eine signifikante Erhöhung der Alpha-Spindelrate im Aggregationsbereich vier ($p < .01$, einseitig getestet).

Darüber hinaus wurde die deskriptiv festzustellende Reduktion der Alpha-Spindelrate zwischen Aggregationsbereich sechs und acht näher geprüft. Da der einzige Automatisierungsfehler in der hohen Zuverlässigkeitsbedingung in Aggregationsbereich sieben präsentiert wurde, sollte wie in Abschnitt 3.1.4.3 erläutert ist, eine etwaige Auswirkung vornehmlich in Aggregationsbereich acht festzustellen sein. Post hoc Vergleiche weisen darauf hin, dass die Alpha-Spindelrate im Aggregationsbereich sieben nicht signifikant niedriger war als im Aggregationsbereich sechs ($p = .09$, einseitig getestet). Allerdings war im Aggregationsbereich acht eine signifikant niedrigere Alpha-Spindelrate als im Aggregationsbereich sechs und sieben zu messen (beide $p < .05$, einseitig getestet).

Für die „ermüdeten“ Probanden aus der niedrigen Zuverlässigkeitsbedingung ist bei Betrachtung der Zeitreihe ebenfalls ein Anstieg der Alpha-Spindelrate zu Beginn der Messung zu beobachten. Deskriptiv endet der Anstieg jedoch bereits in Aggregationsbereich drei und nicht wie in der hohen Zuverlässigkeitsbedingung in Aggregationsbereich vier. Dementsprechend wurde der Anstieg der Alpha-Spindelrate zwischen Bereich eins und drei analysiert. Post hoc Vergleiche weisen darauf hin, dass in Aggregationsbereich zwei ein signifikanter Anstieg der Alpha-Spindelrate im Vergleich zu Bereich eins zu verzeichnen ist ($p < .01$, einseitig getestet). Auch in Aggregationsbereich drei herrscht eine verglichen mit Bereich zwei signifikant angestiegene Alpha-Spindelrate vor ($p < .01$, einseitig getestet). Zwischen Aggregationsbereich drei und acht verhält sich die Alpha-Spindelrate deskriptiv konstant, wobei in Aggregationsbereich sechs eine leichte Reduktion der Rate erkennbar ist. Daher wurde für die Bereiche drei bis acht eine separate Analyse durchgeführt: Kein Paarvergleich unterschreitet das Alphafehlerniveau von .05.

Es wurde die Frage aufgeworfen, welche Faktoren ursächlich für die großen interindividuellen Unterschiede im physiologisch erfassten Ermüdungsverhalten sein könnten. Betrachtet man die häufig berichteten Einflussgrößen für Ermüdung (vgl. Abschnitte 2.3 und 2.4.5), stechen hier der Ermüdungszustand vor dem Experiment, die vorherige Schlafdauer, die Aufnahme von Koffein und das Alter der Versuchsteilnehmer hervor. Daher wurde aufbauend auf der Analyse der intraindividuellen Varianz, die interindividuelle Varianz im Ermüdungsverhalten der „ermüdet“ klassifizierten Probanden untersucht. Hierfür wurden die zuvor definierten acht Aggre-

gationsbereiche beider Zuverlässigkeitsbedingungen verwendet (mit Ausnahme des ersten Bereichs, da hier aufgrund der Normierung für alle Probanden der Messwert null vorlag). Abbildung 42 zeigt zunächst die mittlere relative Alpha-Spindelrate aller „ermüdet“ klassifizierten Probanden ($n = 18$) abhängig von ihrem Koffeinkonsum, der dichotom erhoben wurde. Die statistischen Kennwerte der hypothesengeleiteten zufallskritischen Vergleiche über t -Tests für jeden Aggregationsbereich sind in Tabelle 11 aufgelistet. In den meisten Aggregationsbereichen konnte eine signifikant niedrigere Müdigkeit festgestellt werden, wenn Probanden am Morgen vor dem Versuch Koffein konsumiert hatten. Im Bereich sechs wurden das Signifikanzniveau hierfür nur knapp verfehlt. Daher ist davon auszugehen, dass die vorherige Aufnahme von Koffein eine signifikante Reduktion der Alpha-Spindelrate und damit der objektiv erfassten Ermüdung zur Folge hatte.

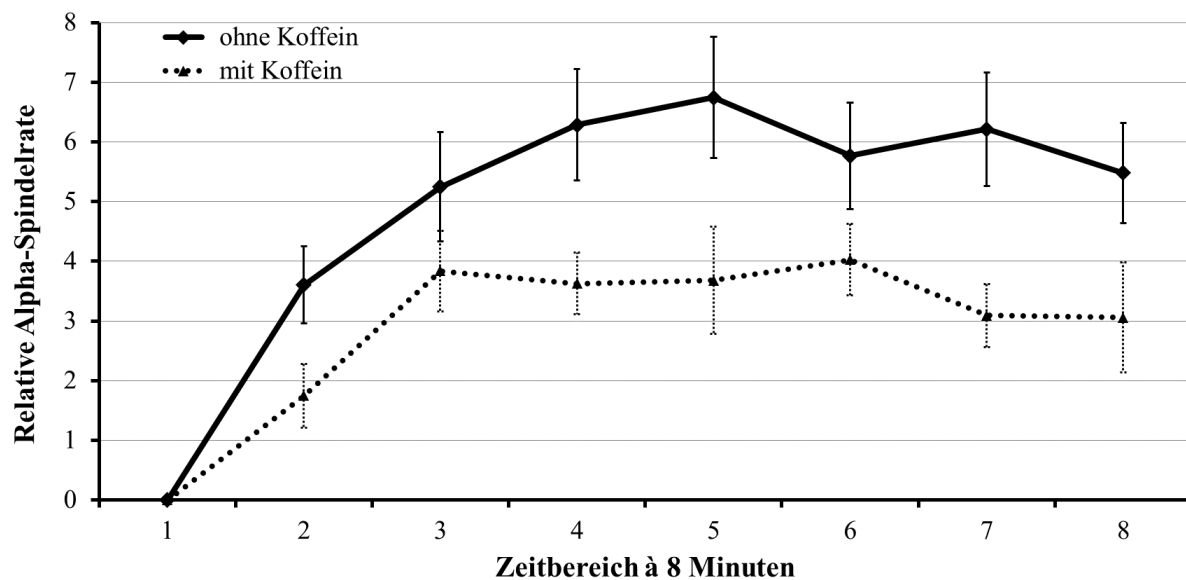


Abbildung 42: Relative EEG-Alpha-Spindelrate für „ermüdet“ klassifizierte Probanden abhängig von der Fahr-
dauer (zeitlich aggregiert) separiert nach Koffeineinnahme. Die Fehlerbalken repräsentieren den
Standardfehler.

Tabelle 11: Statistische Kennwerte von *t*-Tests zwischen Koffeineinnahme und keiner Koffeineinnahme abhängig vom jeweiligen Aggregationsbereich der EEG-Alpha-Spindelrate.

Bereich	2	3	4	5	6	7	8
Minute	8 - 15	16 - 23	24 - 31	32 - 39	40 - 47	48 - 55	ab 56
<i>t</i>	-2.02	-1.11	-2.12	-2.09	-1.62	-2.87	-1.89
<i>df</i>	16	16	16	16	15.74	14.79	16
Effekt*	$p < .05$	$p = .14$	$p < .05$	$p < .05$	$p = .06$	$p < .01$	$p < .05$
η_p^2	.20	.07	.22	.22	.11	.27	.18

* einseitig getestet

Im Gegensatz zum Koffeinkonsum lagen die subjektive Müdigkeit vor dem Versuch, die Schlafdauer und das Alter der Probanden auf Kardinalskalenniveau vor. Dementsprechend konnte der Zusammenhang zwischen diesen beiden Variablen und der mittleren relativen Alpha-Spindelrate für jeden Aggregationsbereich von allen „ermüdet“ klassifizierten Probanden ($n = 18$) bestimmt werden. Tabelle 12 zeigt hierfür die Pearson-Korrelationen. Keine der berechneten Korrelationen konnte zufallskritisch abgesichert werden. Tendenziell – vor allem zu Beginn der Messfahrt in den Aggregationsbereichen zwei und drei – ist zu beobachten, dass mit erhöhter subjektiver Müdigkeit vor dem Versuch eine geringere Alpha-Spindelrate zu beobachten ist. Entgegen der Hypothese zeichnet sich tendenziell in den Aggregationsbereichen vier bis acht eine erhöhte Alpha-Spindelrate mit erhöhter Schlafdauer ab. In der Tendenz haben Probanden mit einer höheren Schlafdauer ihre Müdigkeit vor dem Versuch ebenfalls höher beurteilt ($r = .40$, $p = .10$). Ferner konnte beobachtet werden, dass höhere Korrelationen zwischen dem Alter der Probanden und ihrer mittleren Spindelrate in den Aggregationsbereichen drei bis fünf vorliegen – das Signifikanzniveau wurde teilweise nur knapp verfehlt. Dies kann darauf hindeuten, dass mit zunehmendem Alter die Ermüdungsstärke geringer ausfällt. Zwischen dem Alter der Versuchsteilnehmer und ihrer subjektiven Müdigkeit zu Versuchsbeginn bzw. ihrer vorherigen Schlafdauer in Stunden gab es keine nennenswerten Zusammenhänge.

Tabelle 12: Pearson-Korrelationen zwischen mittlerer relativer EEG-Alpha-Spindelrate und subjektiver Müdigkeit vor dem Fahrversuch sowie der Schlafdauer in der Nacht vor dem Fahrversuch.

Bereich	2	3	4	5	6	7	8
Minute	8 - 15	16 - 23	24 - 31	32 - 39	40 - 47	48 - 55	ab 56
subj. Müd. a priori	$r = -.22$ $p = .39$	$r = -.30$ $p = .23$	$r = -.11$ $p = .67$	$r = -.08$ $p = .76$	$r = -.26$ $p = .30$	$r = -.20$ $p = .43$	$r = .02$ $p = .93$
Schlaf- dauer	$r = -.08$ $p = .76$	$r = -.06$ $p = .81$	$r = .27$ $p = .28$	$r = .36$ $p = .15$	$r = .38$ $p = .13$	$r = .33$ $p = .19$	$r = .26$ $p = .31$
Alter	$r = -.21$ $p = .40$	$r = -.44$ $p = .07$	$r = -.39$ $p = .11$	$r = -.36$ $p = .14$	$r = -.21$ $p = .40$	$r = -.12$ $p = .64$	$r = .07$ $p = .78$

3.2.3.3 Analyse des Einflusses von Automatisierungsfehlern

Den Probanden der niedrigen Funktionszuverlässigkeitsbedingung wurden ungefähr alle acht Minuten ein Automatisierungsfehler präsentiert (vgl. Abschnitt 3.1.1), wobei Probanden, die eine hohe Zuverlässigkeit erlebten, ausschließlich im Bereich von Minute 50 auf eine Querablage reagieren mussten. Um Einflüsse der Automatisierungsfehler auf das Ermüdungsverhalten sichtbar zu machen, wurde die mittlere Alpha-Spindelratendifferenz eine Minute bevor der Fehler in Erscheinung trat und eine Minute nach Auftreten des Fehlers bestimmt. Bezugnehmend auf das Verfahren der Alpha-Spindeldetektion (vgl. Abschnitt 3.1.4.3) wurde ein gleitender Mittelwert der Daten über die Zeit verwendet. Für die vorangegangenen Analyseschritte empfahl sich ein gleitender Mittelwert über 300 Sekunden. Da an dieser Stelle zwei Messzeitpunkte mit einer zeitlichen Differenz von einer Minute vorlagen, hätte der zweite Messzeitpunkt einen Großteil seines Dateneinflusses analog zum ersten Messzeitpunkt aufgewiesen. Aus diesem Grund wurde für diesen Analysefall ein gleitender Mittelwert über 60 Sekunden gewählt. Die Referenzierung wurde analog zum Mittelwert über 300 Sekunden anhand der Spindelrate zu Minute fünf durchgeführt.

Wird ausschließlich der Automatisierungsfehler nach ca. 50 Minuten Level 2 betrachtet, so besteht hier die Möglichkeit, das Ermüdungslevel der Probanden beider Zuverlässigkeitsbedin-

gungen zu vergleichen. Für Probanden, die einer hohen Funktionszuverlässigkeit zugeteilt waren, stellte diese Querablage den ersten (und einzigen) erlebten Automatisierungsfehler dar, für Probanden aus der niedrigen Zuverlässigkeitsbedingung war dies bereits der siebte Fehler. Abbildung 43 zeigt die mittlere relative Alpha-Spindelrate eine Minute vor und eine Minute nach dem Automatisierungsfehler für die Probanden beider Zuverlässigkeitsbedingungen.

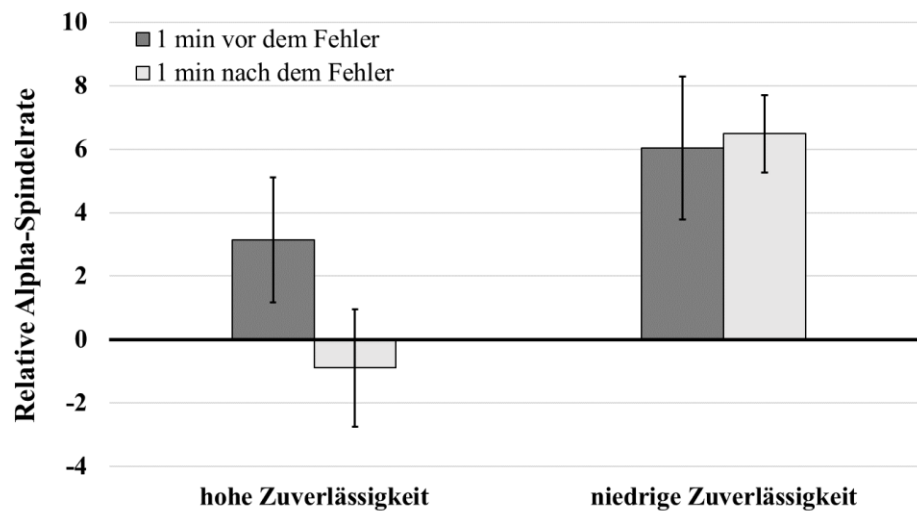


Abbildung 43: Mittlere relative EEG-Alpha-Spindelrate für „ermüdet“ klassifizierte Probanden eine Minute vor und eine Minute nach dem Automatisierungsfehler ungefähr im Bereich von Minute 50, separiert für eine hohe und eine niedrige erlebte Funktionszuverlässigkeit. Die Fehlerbalken repräsentieren den Standardfehler.

Eine 2×2 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf einem Faktor zeigte entgegen der Hypothese keinen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ auf die Ermüdungsstärke ($F(1, 14) = 2.14, p = .17, \eta_p^2 = .13$). Folglich unterschieden sich die hohe und niedrige Zuverlässigkeitsbedingung nicht in der Ermüdungsstärke unabhängig vom Messzeitpunkt (vor bzw. nach dem Automatisierungsfehler). Ebenfalls konnte kein signifikanter Haupteffekt des messwiederholten Faktors „Messzeitpunkt“ auf die Ermüdung unabhängig vor der erlebten Zuverlässigkeit nachgewiesen werden ($F(1, 14) = 3.09, p = .10, \eta_p^2 = .18$). Allerdings konnte eine signifikante Interaktion zwischen erlebter Zuverlässigkeit und Messzeitpunkt festgestellt werden ($F(1, 14) = 13.46, p < .01, \eta_p^2 = .49$). Folglich zeigte sich ausschließlich in der hohen Zuverlässigkeitsbedingung eine Reduktion der Ermüdung nach der erlebten Querablage.

Für die Analyse der Auswirkung von Automatisierungsfehlern in den übrigen sieben Zeitbereichen konnten ausschließlich die Daten von Probanden aus der niedrigen Zuverlässigkeitsbedingung analysiert werden. Hierfür ist die relative Alpha-Spindelrate eine Minute vor dem jeweils erlebten Automatisierungsfehler und eine Minute danach in Abbildung 44 dargestellt. Es wurde für jeden Messzeitpunkt ein t -Test für abhängige Stichproben berechnet. Tabelle 13 zeigt die Ergebnisse der Analyse.

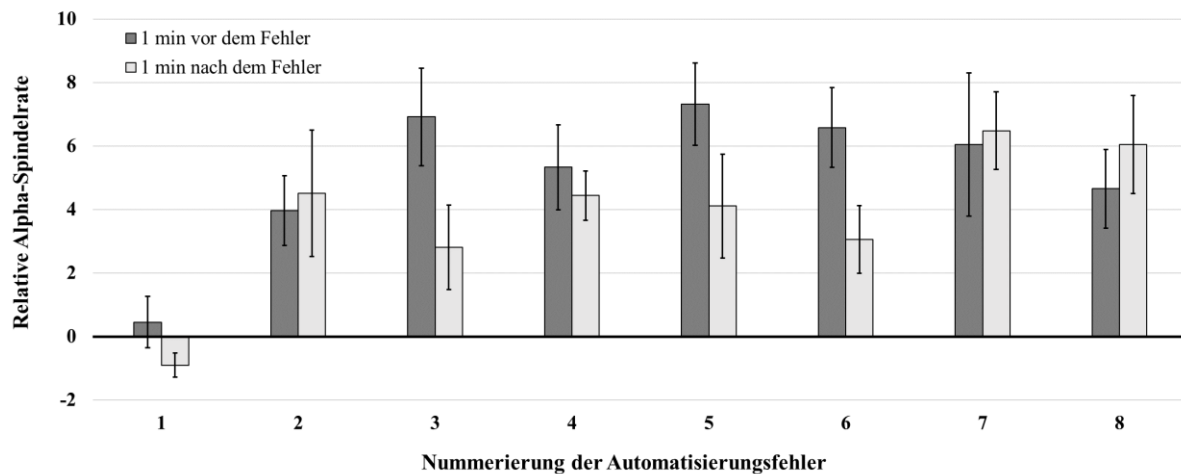


Abbildung 44: Mittlere relative EEG-Alpha-Spindelrate für „ermüdet“ klassifizierte Probanden der niedrigen erlebten Funktionszuverlässigkeit, eine Minute vor und eine Minute nach dem jeweiligen Automatisierungsfehler. Die Fehlerbalken repräsentieren den Standardfehler.

Tabelle 13: Statistische Kennwerte von t -Tests zwischen der EEG-Alpha-Spindelrate eine Minute vor und nach einem Automatisierungsfehler im jeweiligen Aggregationsbereich.

Bereich	1	2	3	4	5	6	7	8
Minute	0 - 7	8 - 15	16 - 23	24 - 31	32 - 39	40 - 47	48 - 55	ab 56
t	-1.53	.29	-2.46	-.58	-3.33	-2.73	.25	.97
df	8	8	8	8	7	7	8	8
Effekt*	$p = .08$	$p = .39$	$p < .05$	$p = .30$	$p < .01$	$p < .05$	$p = .41$	$p = .18$
d	.51	.10	.82	.19	1.18	.96	.08	.32

* einseitig getestet

Für die Bestimmung der Reaktionszeit (Zeit zwischen Beginn des Automatisierungsfehlers und Handtasterdruck des Probanden) wurde der Fehler um Minute 50 herangezogen. Für diesen Fehler lagen die Reaktionszeiten für „ermüdet“ wie „nicht ermüdet“ klassifizierte Probanden sowohl aus der hohen als auch aus der niedrigen Zuverlässigkeitsbedingung vor. Für alle anderen Reaktionen von Probanden auf Automatisierungsfehler bei niedriger Funktionszuverlässigkeit zeichnete sich keine sinnvolle Auswertbarkeit ab: Da unterschiedliche Fehler sowohl auf geraden Teilstücken als auch in den Steilkurven präsentiert wurden, wären keine vergleichbaren Auswertungen in einem Zeitbereich möglich. Insbesondere stellte sich bei der Datenaufbereitung heraus, dass die Sensorkamera deutlich später auf Querablagen in der Steilkurve reagierte, als dies auf geraden Teilstücken der Fall war.

In Abbildung 45 ist die mittlere Reaktionszeit auf die Querablage um Minute 50 für „ermüdete“ und „nicht ermüdete“ Probanden aus beiden Zuverlässigkeitsbedingungen dargestellt. Eine 2×2 faktorielle Varianzanalyse (ANOVA) zeigte einen hypothesenkonformen, signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 28) = 8.23, p < .01, \eta_p^2 = .23$). Demnach reagierten Probanden, die bereits sechs Automatisierungsfehler zuvor erlebt hatten, schneller auf den Fehler als Probanden, bei denen zuvor noch kein Automatisierungsfehler auftrat. Negative mittlere Reaktionszeiten bedeuten hier, dass die Versuchsteilnehmer die Querablage antizipierten, und den Taster vor Überschreitung der Fahrstreifenmarkierung betätigten. Darüber hinaus konnte entsprechend der Hypothese ein signifikanter Haupteffekt des Zwischengruppenfaktors „Ermüdung“ festgestellt werden ($F(1, 28) = 4.17, p < .05, \eta_p^2 = .15$). Insofern konnten schnellere Reaktionen der „nicht ermüdet“ klassifizierten Probanden auf die Querablage gemessen werden. Ein Interaktionseffekt zwischen „Funktionszuverlässigkeit“ und „Ermüdung“ konnte nicht nachgewiesen werden ($F(1, 28) = .23, p = .63, \eta_p^2 = .01$). Außerdem konnte mit Hilfe eines t -Tests für unabhängige Stichproben festgestellt werden, dass das vor dem Fahrversuch konsumierte Koffein keinen Einfluss auf die hier ermittelten Reaktionszeiten hatte ($t(30) = -.52, p = .61, \eta_p^2 = .01$). Zudem existierte kein Zusammenhang zwischen den Reaktionszeiten und dem Probandenalter ($r = -.01, p = .96$).

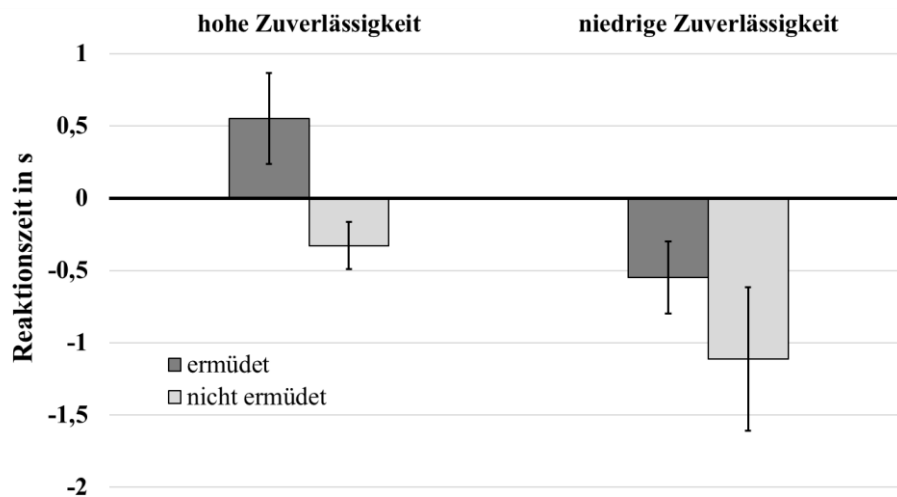


Abbildung 45: Mittlere Reaktionszeit auf den Automatisierungsfehler (Querablage) im Bereich von Minute 50, aufgeteilt auf die hohe und niedrige erlebte Funktionszuverlässigkeit für „ermüdet“ und „nicht ermüdet“ klassifizierte Probanden. Die Fehlerbalken repräsentieren den Standardfehler.

3.2.4 Videobasierte Müdigkeitsbeurteilung

Auf Basis der entsprechend Abschnitt 3.1.7 erklärten Vorgehensweise wurde das von den Versuchsteilnehmern aufgezeichnete Videomaterial in acht ein-minütige Sequenzen aufgeteilt und hinsichtlich spezifischer Müdigkeitsindikatoren von einem geschulten Rater beurteilt. Auf Grund technischer Probleme zeichnen die Videokameras während der Versuchsfahrt von einigen Probanden keine Bilder auf. Des Weiteren erwiesen sich Reflektionen von Sonnenlicht auf Brillengläsern der Probanden als hinderlich für eine fundierte Beurteilung der Müdigkeit, weil hierdurch das Lidschlussverhalten nicht eindeutig erkennbar war. Daher lagen für lediglich $n = 19$ Probanden Müdigkeitsbeurteilungen auf Grundlage der Videoanalyse vor. Hiervon waren neun Personen der hohen erlebten Funktionszuverlässigkeit, zehn Personen der niedrigen Zuverlässigkeit zugeteilt.

Abbildung 46 zeigt die videobasierte Müdigkeitsbeurteilung abhängig vom Messzeitpunkt, der analog zu den acht Aggregationsbereichen der EEG-Daten gewählt wurde. Eine 2×8 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf dem zweiten Faktor zeigte entgegen der Hypothese keinen signifikanten Haupteffekt des Zwischengruppenfaktors „Funktionszuverlässigkeit“ ($F(1, 17) = .01, p = .92, \eta_p^2 < .01$). Folglich unterscheiden sich die Probanden der hohen und niedrigen Zuverlässigkeitsbedingung nicht signifikant in ihrem Ermüdungsverhalten. Betrachtet man jedoch die einzelnen acht Messzeitpunkte, so ist ein signifikanter Haupteffekt entsprechend der Hypothese festzustellen ($F(7, 119) = 7.02, p < .001, \eta_p^2 = .29$). Post hoc

t -Tests zeigten, dass zum zweiten Messzeitpunkt eine signifikant höhere Müdigkeit anhand des Videos beurteilt wurde als zum ersten Messzeitpunkt ($p < .05$). Auch wenn deskriptiv die beurteilte Müdigkeit weiter ansteigt, ist bereits zwischen dem dritten Messzeitpunkt und dem zweiten Messzeitpunkt kein signifikanter Anstieg mehr nachweisbar ($p = .06$, einseitig getestet), auch wenn das Signifikanzniveau nur knapp verpasst wird.

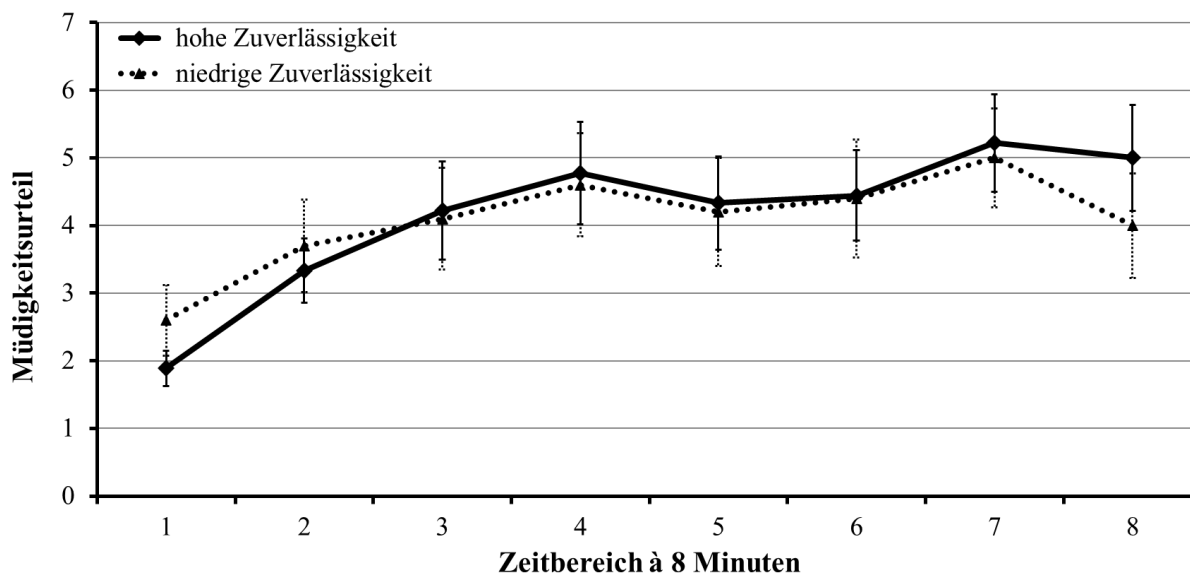


Abbildung 46: Mittlere videobasierte Müdigkeitsurteile abhängig von der Fahrdauer (über acht Messzeitpunkte) für Probanden der hohen und niedrigen erlebten Funktionszuverlässigkeit. Die Fehlerbalken repräsentieren den Standardfehler.

Betrachtet man die Müdigkeitsverläufe auf Grundlage der videobasierten Müdigkeitsbeurteilung abhängig von der auf EEG-Alpha-Spindelbasis klassifizierten Ermüdung, so sind keine signifikanten Unterschiede im Ermüdungsverlauf nachweisbar – vielmehr wird eine hohe Kongruenz beider Verläufe sichtbar, siehe Abbildung 47. Ein spezifischerer Vergleich für alle „ermüdet“ klassifizierten Probanden zwischen ihrer mittleren relativen Alpha-Spindelrate im jeweiligen Aggregationsbereich und ihrer äquivalenten videobasierten Müdigkeitsbeurteilung ($n = 9$) wurde korrelativ durchgeführt. Tabelle 14 zeigt die entsprechenden Pearson-Korrelationen. Demnach ist in den Zeitbereichen zwei, drei und sechs ein signifikanter Zusammenhang der Messgrößen beider Detektionsverfahren erkennbar.

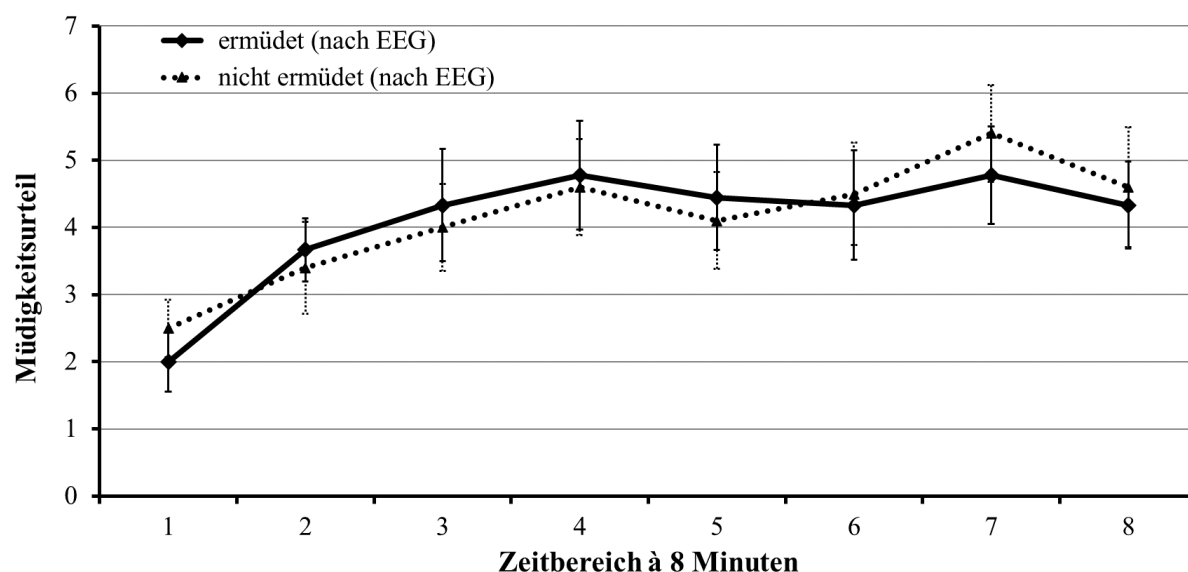


Abbildung 47: Mittlere videobasierte Müdigkeitsurteile abhängig von der Fahrdauer (über acht Messzeitpunkte) für auf Basis ihrer EEG-Alpha-Spindelrate „ermüdet“ oder „nicht ermüdet“ klassifizierter Probanden. Die Fehlerbalken repräsentieren den Standardfehler.

Tabelle 14: Pearson-Korrelationen zwischen mittlerer relativer EEG-Alpha-Spindelrate und videobasierter Müdigkeitsbeurteilung zum jeweiligen Messzeitpunkt.

Bereich	2	3	4	5	6	7	8
Minute	8 - 15	16 - 23	24 - 31	32 - 39	40 - 47	48 - 55	ab 56
Video-Urteil im jeweiligen Bereich*	$r = .82$ $p < .01$	$r = .65$ $p < .05$	$r = .16$ $p = .34$	$r = .04$ $p = .50$	$r = .60$ $p < .05$	$r = .35$ $p = .18$	$r = .13$ $p = .37$

* einseitig getestet

Analog zu der Alpha-Spindel-basierten Müdigkeitsanalyse, wurde der Einflussfaktor des Koffeinkonsums vor dem Fahrversuch auch für die videobasierte Müdigkeitsbeurteilung evaluiert. In Abbildung 48 sind dafür die mittleren Müdigkeitsurteile abhängig vom jeweiligen Messzeitpunkt separiert nach Koffeinkonsum dargestellt. Die statistischen Kennwerte der hypothesengeleiteten zufallskritischen Vergleiche über t -Tests für jeden Zeitbereich sind in Tabelle 15 aufgelistet. Zu einigen Messzeitpunkten konnte eine signifikant niedrigere Müdigkeit auch auf Videobeurteilungs-Basis festgestellt werden, wenn Probanden am Morgen vor dem Versuch Koffein konsumiert hatten. Zu anderen Messzeitpunkten wurde das Signifikanzniveau nur knapp verfehlt.

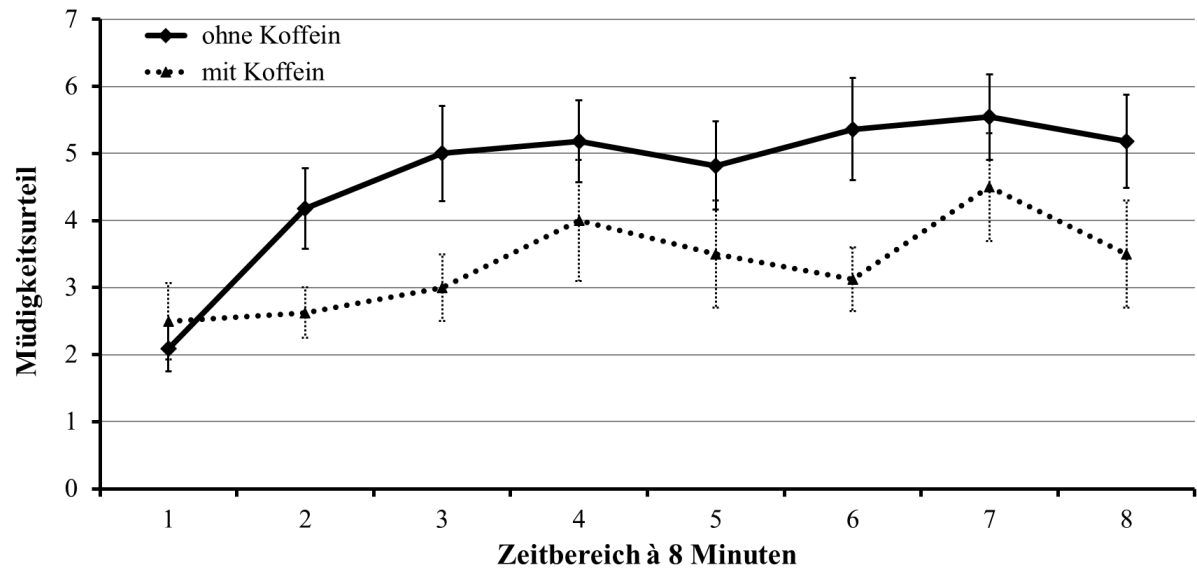


Abbildung 48: Mittlere videobasierte Müdigkeitsurteile abhängig von der Fahrdauer (über acht Messzeitpunkte) separiert nach Koffeineinnahme. Die Fehlerbalken repräsentieren den Standardfehler.

Tabelle 15: Statistische Kennwerte von *t*-Tests zwischen Koffeineinnahme und keiner Koffeineinnahme abhängig vom jeweiligen Messzeitpunkt der Videobeurteilung.

Bereich	1	2	3	4	5	6	7	8
Minute	0 - 7	8 - 15	16 - 23	24 - 31	32 - 39	40 - 47	48 - 55	ab 56
<i>t</i>	.65	-2.20	-2.12	-1.12	-1.28	-2.48	-1.03	-1.58
<i>df</i>	17	15.87	17	17	17	15.89	17	17
Effekt*	<i>p</i> = .26	<i>p</i> < .05	<i>p</i> < .05	<i>p</i> = .14	<i>p</i> = .11	<i>p</i> < .05	<i>p</i> = .16	<i>p</i> = .07
η_p^2	.02	.19	.21	.07	.09	.23	.06	.13

* einseitig getestet

3.3 Beantwortung der Fragestellungen

Mit Verweis auf Abschnitt 2.5.1 wurden insgesamt fünf Forschungsfragen gestellt, die im Rahmen des ersten Experiments beantwortet werden sollten.

1. Die Machbarkeit EEG-basierter Müdigkeitsmessung ist unter Zuhilfenahme der Alpha-Spindelanalyse auch in einer automatisierten Realfahrt gegeben. Auf Grundlage der individuellen Alpha-Spindelraten konnten Nutzer einer Level 2-Funktion entsprechend ihrer gezeigten Ermüdung klassifiziert werden. Dabei verhielten sich hohe Spindelraten proportional zur jeweiligen Müdigkeit, allerdings konnte mit hoher Wahrscheinlichkeit nicht die absolute Müdigkeit über die Fahrdauer gemessen werden, sondern vielmehr das individuelle Ermüdungspotential, das abhängig von der Ausgangsmüdigkeit war. Die intra- und interindividuellen Fluktuationen der Spindelrate waren jedoch immens groß.

Die EEG-basierte Alpha-Spindelanalyse stellt weitestgehend einen validen Müdigkeitsindikator dar. Individuelle Unterschiede im Ermüdungsverhalten konnten äquivalent über die videobasierte Müdigkeitsbeurteilung dargestellt werden – vorausgesetzt man beschränkt sich auf die post hoc „ermüdet“ klassifizierte Teilstichprobe. Es war nicht möglich, die Clusterung in vermeintlich „ermüdete“ und „nicht ermüdete“ Versuchsteilnehmer anhand der videobasierten Müdigkeitsbeurteilung abzubilden. Die Ursache mag darin liegen, dass anhand der Videodaten das absolute Müdigkeitsniveau erfasst wurde, die Alpha-Spindeln jedoch den relativen Müdigkeitszuwachs anzeigten. Nach dem Fahrversuch besaßen die Versuchsteilnehmer subjektiv wie objektiv (anhand der Spindelraten nur die „ermüdet“ klassifizierte Teilstichprobe) eine höhere Müdigkeit als zu Versuchsbeginn.

2. Beim kontinuierlichen Überwachen einer Level 2-Funktion zeigte sich ein im Mittel monotoner Müdigkeitsanstieg über die 15. Fahrminute hinaus, bis etwa zur 20. Minute das Maximalniveau erreicht war. Innerhalb dieser Zeitspanne (es waren keine Aussagen in den ersten fünf Minuten möglich), konnte die Fahrdauer die intraindividuelle Ermüdungsstärke der „ermüdet“ klassifizierten Teilstichprobe mit einer hohen mittleren Varianzaufklärung von $R^2 = .66$ bei 17 von 18 Probanden signifikant vorhersagen. Maßgeblich spielten auch interindividuelle Einflussfaktoren in die Müdigkeitsentwicklung hinein (siehe Beantwortung der fünften Forschungsfrage). Nach 20 Fahrminuten ist eine Müdigkeitssättigung zu beobachten, d.h. die Müdigkeit verweilt auf einem mittleren Niveau. Dieses Muster findet sich ebenfalls in anderen experimentellen Studien.

3. Der gemessene Workload nach Überwachung der Level 2-Funktion lässt sich niedriger beziffern als üblicherweise beim manuellen Fahren zu beobachten ist, liegt jedoch höher als in Level 3 (mit oder ohne fahrfremde Tätigkeiten). Anhand des Stresserlebens konnten ebenfalls keine Anzeichen einer hohen Beanspruchung ermittelt werden, die für die Ermüdung ursächlich gewesen wären. Auch nach der Versuchsfahrt zeichnete sich keine Zunahme im Stresserleben verglichen mit Fahrtbeginn ab. In Summe konnten damit keine Hinweise für eine Überbeanspruchung während Level 2 nachgewiesen werden. Hingegen zeugte ein signifikant geringerer kognitiver Workload der „ermüdet“ klassifizierten Probanden davon, dass vermutlich ein zu niedriges Arousal-Level für den Ermüdungseffekt verantwortlich war. Eine kausale Verknüpfung zwischen höherem Funktionsvertrauen, einer höheren Langeweile und einer dementsprechend höheren Müdigkeit konnten nicht gefunden werden. Insgesamt liefert das erste Experiment erste Hinweise, dass die für Vigilanzaufgaben typische Unterbeanspruchung ihre Anteile zum Ermüdungseffekt beisteuert. Damit können Überbeanspruchungseffekte nicht per se ausgeschlossen werden, da sie ggf. im vorliegenden Experiment nur nicht nachzuweisen waren.
4. Anhand der EEG-Alpha-Spindelrate, der videobasierten Müdigkeitsbeurteilung und der Selbsteinschätzung der Stichprobe konnten innerhalb der ersten ca. 50 Fahrminuten keine Unterschiede im Ermüdungsverlauf abhängig von der Funktionszuverlässigkeit gefunden werden. Versuchsteilnehmer, die nach 50 Minuten den ersten und einzigen Automatisierungsfehler erlebten, wurden durch diesen signifikant in ihrem Fahrerzustand aktiviert. Bei Teilnehmern, die bereits in regelmäßigen Abständen sechs Fehler erlebt hatten, blieb diese Aktivierung aus. Es konnte gezeigt werden, dass die Müdigkeit erst ein höheres Niveau erreicht haben muss, bevor sich der Aktivierungseffekt manifestieren kann – im konkreten Fall musste das Müdigkeitsmaximum nach 20 Fahrminuten hierfür erreicht worden sein. Dann sind Fehlerpräsentationen in der Lage, eine Aktivierung, die ca. zwei Minuten Bestand hat, herbeizuführen. Wenn auf gesättigtem Müdigkeitsniveau regelmäßig Automatisierungsfehler präsentiert werden, verlieren diese ihren aktivierenden Charakter mit zunehmender Häufigkeit – eine Habituation stellt sich ein.

Weder ein höherer Workload verbunden mit höherem Stress (i.S. einer Überbeanspruchung) noch ein höheres Funktionsvertrauen in Verbindung mit höherer Langeweile (i.S. einer Unterbeanspruchung) war bei einer höheren Funktionszuverlässigkeit im Vergleich zu einer niedrigen Zuverlässigkeit zu beobachten. Auch eine niedrigere Steuerungsgüte des Wizard-Fahrers konnte keine niedrigere Ermüdung bewirken. Betrachtet man die Reaktion auf die Querablage nach etwa 50 Minuten, so antizipierten „nicht ermüdet“ klassifizierte

Probanden den Fehler, wohingegen „ermüdete“ Probanden verzögert, nachdem die Fahrstreifenmarkierung überschritten war, reagierten. Bei niedrigerer Funktionszuverlässigkeit, d.h. einer höheren vorherigen Fehlerrate, erfolgen ebenfalls schnellere Reaktionen aufgrund eines Lerneffekts. Es zeichnete sich die Tendenz ab, dass trotz geringer Zuverlässigkeit diese Funktion bereitwilliger genutzt werden würde als eine sehr zuverlässige Funktion, wobei sich der empfundene Komfort zwischen beiden Zuverlässigkeiten nicht unterschied.

5. Tatsächlich konnten erwartungsgemäß immense individuelle Unterschiede im Ermüdungsverhalten aufgefunden werden. Zwei entscheidende Faktoren waren maßgeblich hierfür verantwortlich, die Koffeineinnahme sowie das Alter der Teilnehmer. Ältere Probanden und Probanden, die vor dem Versuch Koffein konsumierten, zeigten eine signifikant geringere Ermüdung über die Level 2-Fahrt hinweg. Mit höherer Anzahl an Schlafstunden, weil damit unter Umständen die Schlafqualität niedrig war, konnte eine höhere Müdigkeit im Sättigungsbereich festgestellt werden. Weder eine höhere Neigung zu Kinetose, noch ein negatives Befinden nach der Level 2-Fahrt bedingten eine stärkere Ermüdung.

3.4 Diskussion

Grundlegendes Ziel des ersten Experiments war es, die Machbarkeit von Müdigkeitsmessung mittels EEG in einer automatisierten Realfahrt zu evaluieren – daher wird dieser Aspekt in Bezug auf die Alpha-Spindelraten-basierte Ermüdungsklassifikation vorab erörtert. Hierauf aufbauend soll diskutiert werden, wie Müdigkeit beim Überwachen einer Level 2-Funktion entsteht und über eine Fahrdauer von etwa einer Stunde verläuft. Die Erfassung intra- und interindividueller Unterschiede im Ermüdungsverhalten beanspruchte nicht nur Limitationen der Ermüdungsmessung während Level 2 aufzudecken, sondern auch als Validitätsaußenkriterium zu fungieren. Im weiteren Diskussionsverlauf sollen die Ergebnisse spezifischer im Hinblick auf die Validität der EEG-Alpha-Spindelrate, der videobasierten Müdigkeitsbeurteilung und der Selbsteinschätzung der Stichprobe verglichen und beurteilt werden. Daran anschließend wird der Einfluss der Level 2-Zuverlässigkeit auf das Ermüdungsverhalten evaluiert, bevor die Übertragbarkeit der Ergebnisse auf die Bearbeitung klassischer Vigilanzaufgaben bewertet wird. Den Abschluss der Diskussion bildet die Rekapitulation des Untersuchungsansatzes aus dem ersten Experiment mit einem Ausblick auf das nachfolgende, zweite Experiment.

Machbarkeit von Müdigkeitsmessung mittels EEG in einer realen Level 2-Fahrt

Die Evaluation der Machbarkeit ist insofern von großer Bedeutung, als dass bereits bestehende Forschungsliteratur auf eine potentielle Ermüdung von Nutzern bei Verwendung einer Level 2-Funktion schließen lässt – allerdings wurden diese Erkenntnisse nicht auf Basis von EEG-Daten zusammengetragen (vgl. Abschnitt 2.4.5.1). EEG-basierte Verfahren, wie z.B. die Alpha-Spindel-Analyse, erzielen jedoch in Bezug auf ihre Validität bei Müdigkeitsmessung im manuellen Fahrkontext eine breite Fürsprache, nicht selten wird ihnen sogar das Attribut des „Goldstandards“ zugeschrieben (vgl. Abschnitt 2.3.5.4). Dass im automatisierten Fahrkontext bis dato auf ihren Einsatz verzichtet wurde, mag versuchsökonomische oder vermeintlich „serienferne“ Gründe haben. Schließlich handelt es sich bei einer EEG-basierten Analyse um eine experimentelle Methodik, deren Anwendung mit größerem Untersuchungsaufwand (speziell im sich bewegenden Realfahrzeug) verbunden ist als beispielsweise subjektive Müdigkeitsabfragen vorzunehmen.

Anhand genau dieser subjektiven Selbsteinschätzung wurde im ersten Experiment deutlich, dass die Stichprobe nach der Versuchsfahrt eine signifikant höhere Müdigkeit aufwies als vor der Fahrt. Die Effektstärke ist mit einem η_p^2 von .85 besonders hoch, die mittlere Müdigkeitszunahme von „wenig schläfrig (–)“ bis hin zu „stark schläfrig (–)“ beträgt ca. sechs Skalenpunkte auf der SOMS. Aus subjektiver Sicht der Probanden war damit die Induktion von Müdigkeit, also die experimentelle Manipulation im monotonen Szenario, erfolgreich.

Nur wenige Probanden mussten aufgrund verrauschter EEG-Daten aus der Auswertung ausgeschlossen werden. Die erste Herausforderung bestand im Zuge der Alpha-Spindel-Analyse vielmehr darin, vermeintlich (objektiv) ermüdete von (objektiv) nicht ermüdeten, bzw. nur geringfügig ermüdeten Probanden zu trennen, um eine sinnvolle Validierungsgrundlage zu schaffen. Obwohl subjektiv alle Probanden ermüdeten, konnte objektiv mitnichten davon ausgegangen werden. Es zeigte sich früh, dass lediglich eine Clusterung in zwei grobe Kategorien sinnvoll – vor allem aber zwingend notwendig war. Die immense Varianz im intra- und interindividuellen Alpha-Spindelverlauf über die Fahrdauer konnte auf diese Weise besser beherrscht werden, um die weiteren Forschungsfragen fundiert beantworten zu können. Bei dieser Vorgehensweise sind jedoch zwei Einschränkungen zu berücksichtigen: Es muss einerseits hinterfragt werden, ob Probanden, die als „nicht ermüdet“ klassifiziert wurden, tatsächlich keine Anzeichen erhöhter Müdigkeit zeigten, oder diese EEG-basiert nur nicht festzustellen waren. Diese grundsätzli-

che Problematik findet sich bereits bei Weinbeer et al. (2017). Andererseits musste ein Kriterium anhand zweier Hochpassfilter definiert werden, deren Anwendung im Resultat annähernd die Hälfte der Stichprobe als „nicht ermüdet“ einordnete – war dies verhältnismäßig?

Das Phänomen des „low voltage EEG“ (Anokhin et al., 1992; Bazanova & Vernon, 2014) bei spezifischen Personen ist bekannt, kann aber nicht als alleinige Lösung dienen, um den hohen Anteil „nicht ermüdeter“ Personen erklären zu können. Betrachtet man die individuellen Verläufe dieser Probanden, so fallen tatsächlich Nulllinien bei manchen Versuchsteilnehmern auf. Die meisten Spindelraten fluktuieren jedoch um null, \pm zwei oder mehr Spindeln, mit erheblichem zeitlichen Anteil kleiner null. Hier lässt sich nicht mehr von „low voltage EEG“ sprechen. Daher müssen andere Variablen für einen Erklärungsversuch herangezogen werden. Als Indiz dafür, dass sehr wohl Müdigkeitsanzeichen bei diesen Personen vorhanden waren, dient die videobasierte Müdigkeitsbeurteilung. Legt man die Müdigkeitsurteile der EEG-basiert als „ermüdet“ und der als „nicht ermüdet“ klassifizierten Versuchsteilnehmer übereinander, wird eine annähernd perfekte Kongruenz sichtbar. Das hieße, dass eine Ermüdung „nicht ermüdet“ klassifizierter Probanden über die Videoanalyse sichtbar wird, die in den EEG-Alpha-Spindeln verborgen bleibt.

Als alternative Überlegung könnte diskutiert werden, ob die Hochpassfilter zu konservativ ausgelegt wurden. Umgekehrt wäre dann zu vermuten, dass die videobasierte Müdigkeitsbeurteilung zu sensitiv arbeitet, und falsche Alarme in Bezug auf ein erhöhtes Müdigkeitslevel liefert. Diese Annahme wird durch – teils statistisch signifikante – Unterschiede anderer Messgrößen zwischen EEG-basiert „ermüdet“ und „nicht ermüdet“ klassifizierter Probanden gestützt. Hierdurch ließe sich auch eine Verhältnismäßigkeit der Filterkriterien ableiten: „Ermüdete“ Versuchsteilnehmer reagierten signifikant langsamer auf die Querablage um Minute 50 als „nicht ermüdete“ Versuchsteilnehmer. Darüber hinaus vermerkten „ermüdete“ Probanden einen signifikant niedrigeren globalen Workload, konkret eine signifikant niedrigere *kognitive* Beanspruchung nach dem Fahrversuch, im Vergleich zu „nicht ermüdeten“ Probanden. Nur knapp konnte eine Abhängigkeit zwischen der Spindel-basierten Klassifikation und vorherigem Koffeinkonsum statistisch nicht nachgewiesen werden – tendenziell wurden jedoch mehr Versuchsteilnehmer als „nicht ermüdet“ klassifiziert, die im Vorfeld Koffein zu sich nahmen. Obwohl die Interaktion zwischen EEG-basierter Ermüdungsklassifikation und dem Faktor „Messzeitpunkt“ (also vor vs. nach der Versuchsfahrt) in Bezug auf die subjektiv beurteilte Ermüdung zufallskritisch nicht abzusichern ist ($p = .11$, zweiseitig getestet), lohnt eine Diskussion der sich abzeichnenden Tendenz: Es könnte vermutet werden, dass Probanden, die bereits vor der Fahrt eine höhere Müdigkeit besaßen, über die Fahrdauer hinweg eine geringere Müdigkeitszunahme

aufweisen, und somit auf Basis ihrer EEG-Alpha-Spindelrate keinen deutlichen Müdigkeitsanstieg zeigen. Dieser Erklärungsansatz besitzt deshalb Evidenz, weil die probandenindividuelle Referenzierung über die jeweilige Alpha-Spindelrate zu ihrem „Ausgangswert“ zur fünften Fahrminute erfolgte. Zu diesem Zeitpunkt kann eine hohe oder niedrige Müdigkeit des jeweiligen Probanden bereits vorgelegen haben (selbst die absolute Spindelrate gibt keinen Aufschluss hierüber). Folglich zeigen die weiteren Verläufe über die Zeit nur noch den relativen Anstieg in Bezug zum Ausgangswert an. Die subjektiv erhobene Müdigkeit könnte hier somit wertvolle Zusatzinformation geliefert haben, in dem Sinne, dass der Fahrer eine hohe Wachheit zu Versuchsbeginn aufweisen musste, um überhaupt nennenswert ermüden zu können. Aus dieser Perspektive würde die EEG-basierte Clusterung dann indirekt das „ausgeschöpfte Ermüdungspotential“ der Versuchsteilnehmer abbilden, wenn diese in „ermüdet“ und „nicht ermüdet“ klassifiziert werden. Dies kann erklären, warum die zuvor thematisierten Unterschiede zur videobasierten Müdigkeitsbeurteilung in Erscheinung traten. Andere abhängige Variablen, wie das Vertrauen in die Level 2-Funktion, der empfundene Stress, die Nutzungsbereitschaft und der empfundene Komfort, aber auch die hervorgerufene Kinetose zeigten keine signifikanten Unterschiede zwischen den Clustergruppen an. Insofern lohnt es, sich mit weiteren Indizien in Bezug auf die Validität der EEG-basierten Ermüdungsdetektion über Alpha-Spindeln auseinanderzusetzen.

Müdigkeitsverlauf in Level 2 unter Berücksichtigung individueller Einflussfaktoren

Werden ausschließlich die Spindelverläufe der „ermüdet“ klassifizierten Probanden betrachtet, zeigt sich ein kontinuierlicher, quasi-linearer Anstieg der Alpha-Spindelrate innerhalb der ersten 20 Minuten Level 2-Fahrt. Leider lässt sich für die ersten fünf Minuten keine Aussage treffen, da aufgrund des verwendeten gleitenden Mittelwerts keine entsprechenden Werte vorlagen. Nach der 20. Minute ist gemittelt über alle Probanden eine Art Sättigung der Spindelrate – und damit der Müdigkeit – zu beobachten. Dieses „Sättigungs-Phänomen“ beschrieben bereits Kündinger et al. (2020) sowie Schömig et al. (2015), der vorherige Ermüdungsanstieg passt sehr gut zu den üblicherweise prognostizierten Vigilanzeinbußen nach einer vergleichbaren Zeitdauer (Parasuraman & Davies, 1976; Pigeau et al., 1995). Generell ist es wichtig zu beachten, dass diese allgemeinen Beobachtungen grundsätzlich von den gemittelten Stichprobendaten abzuleiten sind; auf die Ebene der einzelnen Probanden kann dies nur bedingt übertragen werden, da die Fluktuationen zwischen den minütigen Messzeitpunkten zu groß sind. Hier konnte aber gezeigt werden, dass die *intraindividuelle* Varianz im Ermüdungsverhalten über die ersten 20 Minuten Level 2 doch insofern beschränkt ist, als dass bei 17 von 18 Probanden die Ermüdungsstärke durch die Fahrdauer mit hoher Varianzaufklärung signifikant vorhergesagt werden

konnte. Die Steuerungsgüte des Wizard-Fahrers konnte hingegen keine zusätzliche Varianzaufklärung liefern.

Neben der intraindividuellen Varianz kann auch die *interindividuelle* Varianz in der EEG-Alpha-Spindelrate (der „ermüdet“ klassifizierten Probanden) Hinweise auf die Validität des Verfahrens geben. Als Außenkriterien standen das Alter der Probanden, ihre Anzahl an Schlafstunden vor der Versuchsfahrt und ihre vorherige Koffeineinnahme zur Verfügung. Zunächst soll der Einfluss des Probandenalters näher betrachtet werden: Es konnten teils hohe negative Korrelationen zwischen dem Alter und der Spindelrate in jedem acht-minütigen Aggregationsbereich nachgewiesen werden. Keine der Korrelationen konnte bei zweiseitiger Testung zufalls-kritisch abgesichert werden – dennoch fällt im Zeitbereich drei (also zwischen der 16. und 23. Fahrminute) auf, dass bei einseitiger Testung dies sehr wohl möglich ist ($r = -.44, p < .05$). In diesem dritten Zeitbereich wird gemittelt über alle Versuchsteilnehmer erstmals das maximale Müdigkeitsniveau erreicht. Aber auch in den beiden darauffolgenden Zeitbereichen bleibt die Korrelation mit $r = -.39$ respektive $r = -.36$ hoch. Demnach war vorliegendes Experiment anhand der gemessenen Alpha-Spindelrate in der Lage, den bereits durch Kunder et al. (2018, 2020) während Level 2 beschriebenen Effekt, wonach ältere Menschen eine abgeschwächte Ermüdung zeigen, valide zu replizieren. Natürlich gilt dies wieder nur für die post hoc „ermüdet“ klassifizierte Teilstichprobe. Da bereits bei Clusterung in vermeintlich „ermüdete“ und „nicht ermüdete“ Probanden die a priori vorherrschende subjektive Müdigkeit ein entscheidendes Maß für das Ermüdungspotential darstellte, wurde diese ebenfalls mit dem Alter der Probanden in Bezug gesetzt. Hierbei zeigte sich jedoch keinerlei Effekt. Jüngere Probanden besaßen zu Versuchsbeginn also nicht zwangsläufig ein niedrigeres (subjektives) Müdigkeitsniveau und konnten deshalb stärker ermüden. In der Konsequenz muss eine Erklärung für den Alterseffekt an anderer Stelle gesucht werden.

Kunder et al. (2018, 2020) beschreiben den Alterseffekt leider mehr, als ihn zu interpretieren. Holte (2018) erklärt in Bezug auf das manuelle Fahren, dass ältere Verkehrsteilnehmer seltener unter Müdigkeitseinfluss Autofahren, jedoch beim Fahren diverse kognitive Einbußen kompensieren müssen. Hier sind Probleme bei selektiven Aufmerksamkeitsprozessen und Schwierigkeiten bei der Inhibition von irrelevanten Umgebungsinformationen zu nennen (Holte, 2018). Genau solche kognitiven Fähigkeiten sind insbesondere bei der Bearbeitung von Vigilanzaufgaben (also auch bei einer Level 2-Nutzung) erforderlich und bedingen den überbeanspruchenden Charakter dieser Tätigkeiten (vgl. Abschnitt 2.2.3.2). Daher könnte eine geringere Ermüdung mit höherem Alter womöglich darauf zurückzuführen sein, dass die notwendigen kognitiven Prozesse nicht mehr auf hohem Niveau ausgeführt werden (können), wodurch

eine wesentliche aufgabenbezogene Ermüdungsursache wegbricht. Dass dennoch eine adäquate Leistung erbracht werden kann, zeigte sich bspw. im ersten Experiment an einer vergleichbaren Reaktionsgeschwindigkeit zwischen jungen und älteren Fahrern auf einen Automatisierungsfehler nach 50 Minuten. Daher sind womöglich Kompensationsprozesse älterer Probanden abgelaufen (welcher Art kann die Studie nicht aufklären) – zusätzlich konnte von einer höheren Fahrerfahrung Älterer profitiert werden. Aber auch der schlafbezogene Müdigkeitsfaktor scheint bei Älteren eine entscheidende Rolle zu spielen – bspw. konnten Dijk, Groeger, Stanley und Deacon (2010) zeigen, dass ältere Menschen tagsüber eine geringere Schläfrigkeit aufweisen als Jüngere. Auf das Yerkes-Dodson-Gesetz angewandt (nach Teigen, 1994), waren Ältere in Summe also in der Lage, ihre Beanspruchung auf einem moderaten Level zu halten, und so einer starken, aufgabenbezogenen Ermüdung zu entgehen.

Sobald sich eine Müdigkeitssättigung einstellte (ab Aggregationsbereich vier, also ab der 24. Fahrminute), fallen hohe positive Korrelationen zwischen der Anzahl der Schlafstunden vor dem Versuch und der Alpha-Spindelrate auf. Diese konnten zufallskritisch nicht abgesichert werden. Dennoch deutet das Muster darauf hin, dass mit mehr Schlafstunden eine höhere Ermüdung einhergehen kann. Diese Feststellung mutet auf den ersten Blick paradox an. Auf den zweiten Blick könnte man aber vermuten, dass das Außenkriterium „Anzahl der Schlafstunden vor dem Versuch“ unglücklich gewählt wurde: Stattdessen wäre die „Schlafqualität“ womöglich ein validerer Prädiktor gewesen, da die Anzahl der benötigten Schlafstunden zum Erreichen einer vergleichbaren Schlafqualität ggf. zwischen den Versuchsteilnehmern variierte, es sogar denkbar ist, dass mehr Schlafstunden notwendig waren, um eine niedrige Schlafqualität zu kompensieren. Eine andere Erklärung des Paradoxons könnte abermals im Ermüdungspotential gefunden werden, das die Alpha-Spindelrate scheinbar mit abbildet. Wenn Probanden vor dem Versuch durch mehr Schlafstunden „ausgeschlafener“ waren, besaßen sie ein größeres Ermüdungspotential, das dann EEG-basiert sichtbar wird. Wird jedoch die subjektive Müdigkeit vor dem Versuch mit der Spindelrate zu jedem Zeitbereich korreliert, wird diese Annahme nur bedingt untermauert, da der korrelative Trend zwar in der erwarteten Richtung verläuft, meistens jedoch zu weit entfernt von einer statistisch fundierten Absicherung liegt.

Bei vielen Müdigkeitsstudien, wie auch in der vorliegenden, war eine Müdigkeitssättigung nach 20 Fahrminuten oder später zu beobachten. Wie ist dieses Muster zu erklären? Zunächst kann weitestgehend ausgeschlossen werden, dass dieses Phänomen auf das jeweils verwendete Messverfahren zurückzuführen ist. Hierfür ist die Studienlage auf Basis verschiedener Indikatoren zu homogen. Eine erste Erklärung könnte im Zusammenspiel zwischen der Fahraufgabe wäh-

rend Level 2 und der im Experiment vorherrschenden Motivation der Versuchsteilnehmer gefunden werden. Die Instruktion sah eine fortwährende Aufmerksamkeit der Probanden vor – dass diese nur bedingt aufrechterhalten werden konnte, zeigt die per se stattfindende Ermüdung. Es kann jedoch vermutet werden, dass Probanden den ihnen auferlegten Anforderungen bestmöglich nachkommen wollten (wie so häufig in experimentalpsychologischen Untersuchungen). Hiervon motiviert konnten die Teilnehmer selbstaktivierendes Verhalten zeigen – dazu gehöre auch, sich kurzzeitig von der Überwachungsaufgabe abzuwenden, um anschließend wieder aufmerksamer sein zu können. Folglich hätten die Teilnehmer versucht, nur eine begrenzte Müdigkeit „zuzulassen“. Videodaten, die auf solche Abwendungen hindeuten, wurden im vorliegenden Experiment jedoch nicht analysiert, stattdessen wird an späterer Stelle das Aktivierungspotential der präsentierten Automatisierungsfehler besprochen. Eine zweite Erklärung (insbesondere für vorliegendes Experiment) könnte im Zirkadianen Rhythmus der Probanden liegen. Einerseits war die Notwendigkeit einer homöostatischen Regulation über Schlaf, bzw. höhere Müdigkeit nicht gegeben (Prozess S des Zwei-Prozess-Modells der Schlafregulation; Borbély, 1982), da die Stichprobe nicht schlafdepriviert war. Andererseits sollte zu der Versuchszeit (zwischen 12:00 Uhr und 13:30 Uhr) entsprechend Prozess C des Modells kein Tiefpunkt über den internen Zeitgebermechanismus vorgelegen haben. Zudem müsste sich unabhängig vom jeweiligen Chronotyp der Probanden, zu dieser Zeit kein niedriger Affekt bemerkbar gemacht haben. In der Konsequenz wäre es sehr schwierig gewesen, Probanden über ein mittleres Müdigkeitsniveau heraus zu ermüden. Diese Annahmen besitzen in Summe einen hohen spekulativen Charakter – hier wäre Forschungsbedarf vonnöten, der sich gezielt der Sättigungsfragestellung widmet.

Über die beiden zuvor diskutierten interindividuellen Einflussgrößen hinaus scheint die Einnahme von Koffein die Ermüdung der Teilstichprobe am stärksten beeinflusst zu haben. Die in Abschnitt 2.3.3.2 beschriebenen, aktivierenden Effekte von Koffein konnten a) während der Level 2-Fahrt nachgewiesen werden und b) war es möglich, anhand der EEG-Alpha-Spindelrate zwischen Probanden mit und ohne vorherige Koffeinaufnahme zu differenzieren. Bei Probanden, die Koffein konsumierten, stieg die Müdigkeit langsamer innerhalb der ersten drei Aggregationsbereiche an und verbleibt anschließend auf einem relativ konstanten Niveau im Bereich von vier Spindeln pro Minute. Ohne Koffeinzufuhr stieg die Müdigkeit im Mittel bis in den fünften Aggregationsbereich hinein quasi-exponentiell an und pendelte sich schließlich mit leicht abfallender Tendenz im Bereich von sechs Spindeln pro Minute ein. Zu allen Messzeitpunkten konnte eine signifikant (oder beinahe signifikant) niedrigere Müdigkeit unter Koffeinzufuhr festgestellt werden als bei Versuchsteilnehmern, die beispielsweise keinen Kaffee am Morgen tranken. Hieße das in der Konsequenz, dass eine höhere Vigilanzleistung aufgrund

geringerer Ermüdung, und damit eine sicherere Verwendung einer Level 2-Funktion unter Koffeineinfluss möglich wäre? Solch eine Aussage sollte hier nicht pauschalisiert werden. Schließlich ist nicht nur die objektiv gemessene Müdigkeit als Indikator relevant, sondern auch die (individuelle) Leistung, die der Fahrer im konkreten Zustand zeigt. So sollte man beispielsweise berücksichtigen, dass das zuvor konsumierte Koffein keinen Reaktionsvorteil auf ein kritisches Automatisierungsfehler nach etwa 50 Minuten Fahrdauer brachte. Nachteile von Koffein in Bezug auf die Ermüdung oder eine schlechtere Fahrerleistung lassen sich andersherum jedoch auch nicht argumentieren. Betrachtet man aber die praktische Relevanz der Thematik, so würde es aus Perspektive der Verkehrssicherheit keine alleinige oder gar eine absurde Lösung darstellen, Level 2-Nutzern den Konsum von Koffein nahezu legen – schließlich sollte eine Level 2-Funktion auch ohne die Einnahme zusätzlicher Stimulantien sicher auf Grundlage eines adäquaten Fahrerzustands verwendet werden können.

Vergleich zwischen Alpha-Spindeln, videobasiertem Urteil und Selbsteinschätzung

An dieser Stelle ist es hilfreich, sich mit der konvergenten Validität zwischen EEG-basierter Müdigkeitserfassung und der videobasierten Müdigkeitsbeurteilung weiter auseinanderzusetzen. Wie zuvor thematisiert, war die Müdigkeitsbeurteilung zwar nicht in der Lage, zwischen der EEG-basierten Ermüdungsklassifikation zu differenzieren, allerdings vermag die Methodik in vergleichbarer Qualität zur Analyse der Alpha-Spindelrate, den aktivierenden Charakter von Koffein abzubilden. Trotzdem muss ein leicht veränderter Ermüdungsverlauf verglichen mit dem EEG berücksichtigt werden: Unter Koffeineinfluss zeigte sich über alle Messzeitpunkte ein moderater Anstieg der Müdigkeit (um einen bis eineinhalb Skalenpunkte). Der Müdigkeitsverlauf schlug sich hingegen im EEG unter Koffeineinfluss – insbesondere in den ersten drei Aggregationsbereichen – deutlich stärker nieder. Für Probanden ohne Koffeinaufnahme ist der quasi-exponentielle Anstieg mit anschließender Sättigung, die videobasiert früher (d.h. bereits nach dem dritten Messzeitpunkt) in Erscheinung tritt, annähernd vergleichbar zur Alpha-Spindelrate.

Generell kann der Alpha-Spindel-Analyse (wenn ausschließlich die „ermüdet“ klassifizierte Teilstichprobe betrachtet wird) eine ausreichende Validität angelehnt an die videobasierte Müdigkeitsbeurteilung bescheinigt werden. Dies fällt insbesondere an den Merkmalen „Müdigkeitsanstieg“ und „Müdigkeitssättigung“ auf. Für beide Merkmale ist eine hohe Kongruenz zwischen den Müdigkeitsdetektionsverfahren zu beobachten. Einen zentralen Müdigkeitsindikator innerhalb der videobasierten Beurteilung stellte die Lidschlussdauer dar – definitionsgemäß mussten Lidschlüsse von mindestens einer halben Sekunde auftreten, damit nach ca.

20 Minuten Fahrt die Stufe fünf „mittlere Müdigkeit (+)“ erreicht werden konnte. Dieses Ergebnis geht mit der Beobachtung von Körber et al. (2015) einher, wonach in den ersten 15 Minuten Level 2 ein signifikanter Anstieg der Lidschlussdauer beobachtet werden konnte – in den anschließenden 25 Minuten jedoch nicht mehr. Demensprechend sind die Ergebnisse der Studie von Körber et al. (2015) teilweise auch mit der in vorliegender Arbeit beobachteten Müdigkeitssättigung vereinbar. Allerdings beobachteten Körber et al. (2015) im weiteren Fahrtverlauf zusätzliche Hinweise auf einen Müdigkeitsanstieg (z.B. anhand einer signifikanten Reduktion des Pupillendurchmessers bis ca. Minute 25). Diesen Müdigkeitsindikator bezog die hier verwendete videobasierte Müdigkeitsbeurteilung aber nicht mit ein, sodass diesbezüglich keine Vergleiche möglich sind.

Validitätsabstriche müssen bei der videobasierten Beurteilung hinsichtlich der Sensitivität auf präsentierte Automatisierungsfehler gemacht werden – eine genauere Diskussion hierzu folgt, sobald die Auswirkung der Level 2-Zuverlässigkeit auf das Ermüdungsverhalten thematisiert wird. Die Müdigkeitsurteile anhand der Videodaten erlauben ferner eine Quantifizierung der Ermüdungsstärke. Wo die relative Alpha-Spindelrate nicht in der Lage ist, das Müdigkeitskonstrukt absolut (und mittels Verbalankern) zu benennen, kann das Rating hier Abhilfe schaffen: Entsprechend der Bezeichnung, die in MUED 8.1 verankert ist (Vogelpohl et al., 2017), steigt die Müdigkeit von „leichte Müdigkeit (–)“ auf ein Niveau zwischen „mittlere Müdigkeit (0)“ und „mittlere Müdigkeit (+)“ an. Mit Koffeinzufuhr beläuft sich das Maximum meist auf „mittlere Müdigkeit (–)“ bis „mittlere Müdigkeit (0)“, ohne Koffein auf „mittlere Müdigkeit (+)“ bis „starke Müdigkeit (–)“. Interessant ist, dass die subjektive Müdigkeitsbewertung (über den subjektiven Teil der SOMS) gut mit der objektiven Müdigkeitsbewertung anhand der Videodaten übereinstimmt. Vor dem Fahrversuch ist sowohl die subjektive wie auch die objektive Müdigkeit auf einem vergleichbaren, geringen Niveau. Nach der Fahrt bewegen sich beide Beurteilungsmaße auf einem oberen, mittleren Niveau. In diesem Fall schienen die Versuchsteilnehmer ihre Müdigkeit nach dem Versuch sehr konsistent zur (vermeintlich) tatsächlich vorherrschenden Ermüdung einzuschätzen.

Diese „Harmonie“ zwischen subjektiver und objektiver Müdigkeitsbeurteilung steht den Beobachtungen von Schmidt et al. (2009) anhand der KSS konträr gegenüber. Eine Erklärung könnte zunächst im experimentellen Aufbau des Versuchs gefunden werden: Die Versuchsteilnehmer von Schmidt et al. (2009) fuhren manuell im öffentlichen Straßenverkehr. Im vorliegenden Experiment fand eine automatisierte Fahrt gemäß SAE Level 2 auf einer Teststrecke statt. Obwohl die Fahrdauer bei Schmidt et al. (2009) deutlich länger war (ca. vier Stunden), war a) die Fahraufgabe eine andere und b) könnte die hohe Monotonie der Teststrecke eine

höhere subjektive Ermüdung suggeriert haben. Auch die Arbeitsgruppe um Kunderling hat die subjektive Müdigkeit während Level 2 im Fahrsimulator (Kunderling et al., 2018) als auch im Realfahrzeug auf der Teststrecke (Kunderling et al., 2020) untersucht. In beiden Studien liegt das maximal erreichte Müdigkeitsniveau knapp vor dem Skalenwert sechs („einige Anzeichen von Müdigkeit“) auf der KSS. Im Fahrsimulator steigt die Müdigkeit bis zu diesem Maximum über die 45 Minuten Fahrt kontinuierlich an, unter höherer Monotonie im Realfahrzeug (weil nur 25 km/h Fahrgeschwindigkeit), ist diese bereits nach 35 Minuten erreicht, mit anschließender Sättigung. Verglichen mit den Ergebnissen aus vorliegendem ersten Experiment ist der subjektiv erfasste Anstieg mittels KSS deutlich flacher – nach 20 Minuten zeigen die Ergebnisse von Kunderling et al. (2018, 2020) beispielsweise, dass lediglich ein mittlerer Skalenwert von vier („ziemlich wach“) vorlag. Da beide Studien eine hohe bis sehr hohe Monotonie aufwiesen, wird angenommen, dass die Probanden bereits nach diesen 20 Minuten eine zumindest mittlere Müdigkeit aufgewiesen hätten, wenn die Probanden nicht durch die regelmäßigen Selbstbewertungen gestört und damit aktiviert worden wären. Darauf deutet auch hin, dass das (vermeintliche) Müdigkeitsmaximum erst 15 bis 20 Minuten später eintrat als dies in vorliegender Untersuchung oder auch bei Miller et al. (2015) der Fall war. Da Kunderling et al. (2018, 2020) diese subjektiven Abfragen inmitten des Ermüdungsprozesses einsetzten (selbst wenn diese subtil über Ein- und Ausblendung auf einem Touchdisplay erfolgten), besteht Grund zu der Annahme, dass ihre Ergebnisse eine Müdigkeitsunterschätzung zeigen.

Bis hierher lässt sich das Ermüdungsverhalten während Level 2 und die Validität der EEG-basierten Ermüdungsdetektion über Alpha-Spindeln folgendermaßen beschreiben: Nach dem Fahrversuch konnte subjektiv ein signifikant höheres Müdigkeitsniveau gemessen werden als zu Versuchsbeginn. Objektiv war dies auf Basis der videobasierten Müdigkeitsbeurteilung auf geringerem Niveau ebenfalls möglich, jedoch nicht für alle Probanden anhand der EEG-Alpha-Spindelrate. Für die „ermüdet“ klassifizierte Teilstichprobe zeichnete sich über die Fahrdauer eine erwartungsgemäße Ermüdung ab. Ein rapider, monotoner Müdigkeitsanstieg in den ersten 20 Minuten spiegelt dabei den typischen Anstieg von Vigilanzdekrement bei der Bearbeitung von Vigilanzaufgaben (hier die Level 2-Nutzung) wider. Annähernd die Hälfte der Stichprobe („nicht ermüdet“) besaß jedoch einen undifferenzierten, stark fluktuierenden Spindelraten-Verlauf, der womöglich das Resultat eines geringen Ermüdungspotentials sein kann, da diese Versuchsteilnehmer bereits mit höherer subjektiver Müdigkeit den Versuch begannen. Die intraindividuelle Ermüdungsstärke konnte über die Fahrdauer valide abgebildet werden, insbesondere interindividuelle Faktoren, wie das Probandenalter oder der Koffeinkonsum, konnten erfolgreich interindividuelle Varianz in den Alpha-Spindelraten erklären. Auch die videobasierte

Müdigkeitsbeurteilung konstatiert der Alpha-Spindelanalyse aufgrund vergleichbarer Müdigkeitsverläufe eine ausreichende Validität. Allerdings bleibt unklar, ob die Videoanalyse mit höherer Sensitivität Müdigkeit aufdecken kann als es über die Alpha-Spindelanalyse möglich ist, oder das Müdigkeitsrating aufgrund niedriger Spezifität schlichtweg nicht in der Lage ist, zwischen fahrerindividuellem Ermüdungspotential zu differenzieren, wozu die Alpha-Spindelanalyse anscheinend fähig wäre. Hier wären weitere Untersuchungen notwendig, um die Möglichkeit einer videobasierten Ermüdungspotentialabschätzung zu erforschen.

Auswirkungen der Level 2-Zuverlässigkeit auf das Ermüdungsverhalten

Im nächsten Schritt soll der Diskussionsfokus auf die Rolle der Funktionszuverlässigkeit beim Ermüden über die Fahrdauer gerichtet werden. Entsprechend der Hypothese wurde vermutet, dass Automatisierungsfehler auf den Fahrerzustand aktivierend wirken könnten. Hierfür müssten diese Fehler folglich in der Lage sein, die Monotonie der Vigilanzaufgabe erfolgreich zu durchbrechen. In Experiment 1 vorliegender Arbeit konnten eine hohe und eine niedrige Funktionszuverlässigkeit realisiert werden, indem Abweichungen von der optimalen Steuerungsgüte der Level 2-Funktion sehr selten (einmalig nach ca. 50 Minuten) oder häufiger (d.h. regelmäßig alle ca. acht Minuten) in Erscheinung traten. Um sich der Beantwortung dieser Fragestellung zu nähern, wurde die Teilstichprobe ausgewählt, die entsprechend ihrer Alpha-Spindelrate als „ermüdet“ klassifiziert wurde. Betrachtet man zunächst die zeitlichen Verläufe für jede Minute automatisierter Fahrt, so fällt in der Bedingung „hohe Funktionszuverlässigkeit“ folgendes Muster auf: Die Müdigkeit steigt annähernd monoton bis ca. Minute 30 an, bleibt bis etwa Minute 50 mit Schwankungen auf einem gesättigten Niveau, bis im Anschluss an die erlebte Querablage die Müdigkeit im Mittel auf ein Niveau absinkt, das ungefähr zu Minute 15 vorlag. Ein anderes Muster zeichnete sich in der Bedingung „niedrige Funktionszuverlässigkeit“ ab: Die Müdigkeit steigt schneller, ebenfalls monoton bis zur Minute 18 an. Im weiteren Verlauf sind Müdigkeitssprünge zu beobachten, also ein phasisches Absinken und Ansteigen der Müdigkeit bis zum Fahrtende. Hervorzuheben ist, dass es im direkten Vergleich mit der Bedingung „hohe Funktionszuverlässigkeit“ zu keinem deutlichen Absinken der Müdigkeit nach Minute 50 kommt. Die Phasenlänge der Sprünge beträgt interessanterweise ca. acht Minuten – also gleichermaßen die Frequenz, in der die Automatisierungsfehler in Längsführung und als Querablage im Versuchsdesign verankert waren. Dabei ist nicht fortwährend nach jedem Präsentationszeitpunkt eines Automatisierungsfehlers eine Aktivierung im Fahrerzustand zu beobachten: Die ersten beiden Fehler schienen keine aktivierende Wirkung zu besitzen. Erst mit dem dritten Fehler könnte man mutmaßlich – weil des Öfteren leicht versetzt zum Absinken der Müdigkeit

– auf rein deskriptiver Ebene von einer aktivierenden Wirkung der Automatisierungsfehler ausgehen. Da diese Abschätzung noch sehr unspezifisch ist, soll im weiteren Verlauf der Diskussion hierauf präziser eingegangen werden.

Es ist festzuhalten, dass Nutzer der Level 2-Funktion unabhängig von der Zuverlässigkeit monoton ermüden; dass dies im Mittel etwas langsamer geschieht, wenn keine Automatisierungsfehler in den ersten 30 Minuten Fahrt auftraten, muss keiner Gesetzmäßigkeit unterliegen, zumal die immens hohe interindividuelle Varianz solche Interpretationen limitiert. Im weiteren Fahrtverlauf schienen regelmäßige Fehler die Fahrer kurzzeitig zu aktivieren. In den ersten 20 Minuten besaß die Müdigkeit womöglich ein noch zu geringes Niveau, als dass von einer Durchbrechung der Monotonie profitiert werden konnte. Probanden, die 50 Minuten lang eine nahezu perfekt arbeitende Level 2-Funktion erlebten, schienen maßgeblich von dem aktivierenden Potential der Querablage zu profitieren. Hingegen kann eine Form von Habituation auf die regelmäßigen Fehler in der Bedingung „niedrige Funktionszuverlässigkeit“ gemutmaßt werden, weil selbst mit hoher Müdigkeit nach 50 Minuten Level 2 keine nennenswerte Aktivierung mehr stattfindet – im Gegenteil steigt die Müdigkeit derweil sogar an. Dieses Phänomen beschreiben auch Kreuzmair et al. (2017), deren Übernahmeaufforderungen zum manuellen Fahren so aktivierend auf die Probanden wirkten, dass die hypothetisierten Müdigkeitseffekte hiervon bis zur Unkenntlichkeit überlagert wurden.

Für detailliertere Interpretationen empfiehlt es sich, die EEG-Alpha-Spindelrate der Versuchsteilnehmer eine Minute vor und nach der Präsentation der Automatisierungsfehler zu betrachten. Nach etwa 50 Minuten konnte in der Bedingung „hohe Funktionszuverlässigkeit“ eine signifikante Müdigkeitsreduktion eine Minute nach der Querablage nachgewiesen werden. Dieser Effekt zeichnete sich in der Bedingung „niedrige Funktionszuverlässigkeit“ nicht ab, stattdessen bleibt die Müdigkeit nach dem Fehler in etwa auf dem Niveau wie vor dem Fehler. Insofern konnten die zuvor deskriptiv thematisierten Unterschiede zwischen beiden Funktionszuverlässigkeiten auch zufallskritisch nachgewiesen werden. In der Bedingung „niedrige Funktionszuverlässigkeit“ konnten analog nach den Automatisierungsfehlern drei bis sechs eine Müdigkeitsreduktion (meist statistisch signifikant) festgestellt werden – jedoch nicht nach dem ersten, zweiten, siebten oder achten Fehler. Dieses Muster manifestiert die Vermutung, wonach in den ersten ungefähr 15 Minuten eine zu geringe Müdigkeit vorlag, als dass ein Aktivierungseffekt der Fehler sichtbar werden würde, nach etwa 50 Minuten fand ggf. eine zu starke Gewöhnung an eine regelmäßige Aktivierung statt, weswegen sich auch hier kein Effekt mehr einstellen konnte. Hier ist es sinnvoll, die Ergebnisse von Schömig et al. (2015) in Relation zu betrachten:

Statt Automatisierungsfehler zu präsentieren, forderten Schömig et al. (2015) zu vier verschiedenen Müdigkeitsniveaus ihre Probanden auf, die manuelle Fahrzeugsteuerung zu übernehmen. Hierdurch konnte immer – unabhängig vom generellen Müdigkeitsniveau – eine Zunahme der Aktiviertheit beobachtet werden. Auch wenn diese Ergebnisse zufallskritisch nicht abgesichert werden konnten, scheint eine Transition ein vielfach höheres Aktivierungspotential zu besitzen, als im vorliegenden ersten Experiment durch die präsentierten Automatisierungsfehler realisierbar war.

Dass im vorliegenden Fall von einer Habituation auf regelmäßige Automatisierungsfehler auszugehen ist, wird zusätzlich durch einen sich abzeichnenden Lerneffekt untermauert: Probanden zeigten unabhängig von ihrer EEG-basierten Ermüdungsklassifikation eine um ca. eine Sekunde schnellere Reaktion auf die Querablage nach ca. 50 Minuten, wenn sie bereits sechs Fehlerpräsentationen zuvor erlebt hatten. Hier lohnt eine vergleichende Einordnung der Reaktionszeiten: Versuchsteilnehmer, die erstmalig auf einen Fehler reagieren mussten und als „nicht ermüdet“ klassifiziert wurden, reagierten in etwa so schnell wie Teilnehmer, die bereits sechs Fehler erlebt hatten, jedoch als „ermüdet“ klassifiziert wurden. Bei beiden Probandengruppen lag die Reaktionszeit im Mittel 400 ms vor dem Überschreiten der Fahrstreifenmarkierung. Obwohl die Instruktion besagte, erst *nachdem* das Fahrzeug die Markierung überschritten hatte, den Handtaster zu betätigen, schien eine Antizipation des bevorstehenden Verlassens des eigenen Fahrstreifens stattgefunden zu haben. Diese Antizipation verläuft bei Probanden, die zuvor durch Fehler „trainiert“ waren und als „nicht ermüdet“ klassifiziert wurden, noch schneller ab, nämlich ca. eine Sekunde bevor der Fehler per Definition relevant wurde. Hiervon unterscheiden sich Probanden deutlich, die erstmalig auf die Querlablage reagieren mussten und vermeintlich ermüdet waren. Diese betätigten im Mittel erst ca. 500 ms nachdem die Fahrstreifenmarkierung überschritten wurde den Handtaster – die Differenz zu den „trainierten, nicht ermüdeten“ Probanden beträgt also ca. 1500 ms. Der hier nachgewiesene Reaktionsnachteil aufgrund von Müdigkeit fügt sich gut in Ergebnisse von Neubauer et al. (2014) und Vollrath et al. (2011) ein, wonach eine ACC-Nutzung eine um etwa 500 ms verzögerte Bremsreaktion auf ein einscherendes Fahrzeug bzw. eine um etwa fünf Sekunden verspätete Geschwindigkeitsanpassung bewirkte. Der ermittelte Habituationseffekt ist indes sinnvoll mit dem von Stapel et al. (2017) gezeigten Reaktionsvorteil in der Detection Response Task (DRT) von erfahrenen Level 2-Nutzern in Einklang zu bringen.

Die Faktoren „Ermüdung“ sowie erzieltes „Training“ anhand der Funktionszuverlässigkeit können somit entscheidend bedingen, wie schnell während Level 2 auf ein sicherheitskritisches Fahrverhalten der Fahrfunktion reagiert werden kann. Da der Fahrer aus Level 2 auf solche

fehlerhaften Regelungstätigkeiten der Funktion jederzeit und unmittelbar reagieren muss, wird mit diesem Ergebnis die praktische Relevanz für die langandauernde Nutzung eines Level 2 deutlich: Wenn sich der Fahrer aufgrund der „trügerischen“ Perfektion der funktionalen Regelung „in Sicherheit wiegt“ und gleichzeitig durch die ihm auferlegte Überwachungstätigkeit (in Form einer Vigilanzaufgabe) ermüdet, könnten sicherheitskritische Fahrsituationen entstehen. Im vorliegenden Szenario würde er das schleichende Verlassen des eigenen Fahrstreifens unter Umständen zu spät (hier um bis zu eineinhalb Sekunden verzögert) bemerken. Ob hiermit gleichermaßen eine adäquate Reaktion ausbleiben wird, lässt sich anhand der Ergebnisse aufgrund der experimentellen Konzeption nicht vorhersagen – schließlich handelte es sich in der vorliegenden Studie lediglich um einen Tastendruck, den die Probanden leisten mussten. Eine Auswahl aus verschiedenen Handlungsmöglichkeiten (z.B. Lenken, Bremsen oder die Kombination aus beidem) würde sicherlich andere, ggf. höhere (kognitive) Anforderungen an den Fahrer stellen. Hierfür würde das Situationsbewusstsein des Fahrers dann einen entscheidenden Faktor darstellen. Ein erster Einblick in die Thematik ist der Studie von Gonçalves et al. (2016) zu entnehmen. Hiernach berichten die Autoren von signifikant höheren Querbeschleunigungen bei müden im Vergleich zu nicht müden Probanden.

Wie zuvor diskutiert wurde, kann das Zusammenwirken von Müdigkeit und einer „trügerischen“ Funktionszuverlässigkeit dieses sicherheitskritische Problem hervorrufen. Bedingt womöglich eine hohe Funktionszuverlässigkeit eine höhere Fahrermüdigkeit? Auch diese Annahme sollte im Rahmen des ersten Experiments geprüft werden. Alle drei Müdigkeitsindikatoren (subjektive Müdigkeitsbeurteilung der Probanden, EEG-Alpha-Spindelrate und videobasierte Müdigkeitsbeurteilung) zeigen hier ein homogenes Bild: Es konnten zwischen der hohen und niedrigen erlebten Funktionszuverlässigkeit zu keinem Messzeitpunkt signifikante Unterschiede in der Ermüdung festgestellt werden. Ausschließlich bei Betrachtung der über die acht Aggregationsbereiche aufgetragenen Alpha-Spindelrate (Abbildung 41, vgl. Abschnitt 3.2.3.2) könnte man im letzten Bereich vermuten, dass nach dem Erleben der niedrigen Zuverlässigkeit eine höhere Müdigkeit vorherrscht als nach dem Erleben der hohen Zuverlässigkeit. Dieses Ergebnis deckt sich mit dem zuvor thematisierten Potential des einzig auftretenden Automatisierungsfehlers, nach selbst 50 Minuten Level 2-Fahrt das Aktivitätsniveau wieder zu steigern. Die zufallskritische Absicherung scheitert hier höchstwahrscheinlich an dem experimentellen Design, bzw. an der damit einhergehenden schwächeren Teststärke des verwendeten statistischen Verfahrens: Zwischen den einzelnen Aggregationszeitpunkten (z.B. sechs und acht in der Bedingung „hohe Zuverlässigkeit“) konnte eine signifikante Reduktion der Müdigkeit zum Zeitpunkt acht ermittelt werden. Obwohl das Müdigkeitsniveau in der Bedingung „niedrige Funktionszuverlässigkeit“ zum Aggregationszeitpunkt acht vergleichbar hoch ist, wie zum

Zeitpunkt sechs in der Bedingung „hohe Zuverlässigkeit“, geht Teststärke aufgrund des Zwischengruppenvergleichs verloren, von der im messwiederholten Vergleich profitiert wurde. Die Funktionszuverlässigkeit als messwiederholte Variable zu untersuchen, wäre allerdings – sowohl aufgrund der daraus resultierenden Dauer des Fahrversuchs als auch aufgrund mangelnder Vermittelbarkeit wechselnder Zuverlässigkeiten gegenüber den Probanden (i.S. einer niedrigen externen Validität) – nicht zielführend gewesen.

Bis hierhin kann festgehalten werden, dass sich der Ermüdungsverlauf über die Zeit global gesehen nicht zwischen der Verwendung einer hohen und niedrigen Level 2-Funktionszuverlässigkeit unterscheidet. Vielmehr sind einzelne Automatisierungsfehler in der Lage, kurzzeitig, d.h. für ca. zwei Minuten, eine Aktivierung des Fahrers hervorzurufen. Diese ist jedoch nur dann erfolgreich, wenn der Fahrer a) müde genug ist und b) nicht zu viele Automatisierungsfehler zuvor erlebt hatte. Andernfalls kann ein Trainings- bzw. Habituationseffekt auftreten, sodass die Fehler quasi wirkungslos in Hinblick auf ihren aktivierenden Charakter erscheinen. Wenn Fehler häufiger und regelmäßig auftreten, kann – eben aufgrund des stattfinden Trainings – auch nach langer Fahrdauer (hier 50 Minuten) und unter höherer Müdigkeit, verhältnismäßig schnell auf einen Fehler reagiert werden. Die hier gemessene Reaktionsgeschwindigkeit beinhaltet aber per se keine Aussagen zur Adäquatheit der Reaktionsauswahl.

Übertragbarkeit typischer Merkmale von Vigilanzaufgaben auf eine Level 2-Nutzung

Neben den zuvor diskutierten Zusammenhängen zwischen der Müdigkeitsentwicklung über die Zeit während Level 2, der Validität verschiedener Messverfahren und dem Aktivierungspotential von Automatisierungsfehlern, sollen nun weitere interindividuelle Einflussgrößen auf die Ermüdung der Fahrer betrachtet werden. Es interessierte, ob typische Beanspruchungsmerkmale, die bei der Bearbeitung klassischer Vigilanzaufgaben anzutreffen sind, auch bei Level 2-Nutzung zu beobachten sind. Eine wesentliche Orientierung soll hierfür das in Abschnitt 2.3.4 angepasste Müdigkeitsmodell bieten. Demnach müssten sowohl Merkmale vorliegen, die auf eine Überbeanspruchung als auch auf eine Unterbeanspruchung schließen lassen. Zunächst sollen die Auswirkungen der experimentellen Manipulation anhand des Faktors „Funktionszuverlässigkeit“ diskutiert werden.

Das absolute Workload-Niveau nach Level 2-Nutzung kann im ersten Experiment mit etwa 33 % der mittels NASA-TLX maximal erfassbaren Beanspruchung beziffert werden. In Relation zu dem von De Winter et al. (2014) zusammengetragenen Workload von 43.5 % in Level 0, 38.6 % in Level 1 und 22.7 % in Level 3 (sowohl mit als auch ohne fahrfremde Tätigkeiten),

würde sich der Wert für ein Level 2 prinzipiell plausibel einordnen lassen. Wie ist dies zu interpretieren? Von Überbeanspruchung, insbesondere wenn beim manuellen Fahren über 40 % erreicht werden, sollte hier nicht gesprochen werden. Im Gegenteil sinkt die Beanspruchung unter die in Level 1 bekannte ab. Damit wäre ein Indiz gefunden, wonach im vorliegenden Experiment der gemessene Workload ein Abbild eines niedrigen Arousal-Niveaus ist, welches auf eine Unterbeanspruchung hindeutet. Im direkten Vergleich zwischen beiden konstruierten Funktionszuverlässigkeiten sollte gemäß dem Modell bei hoher Zuverlässigkeit (also seltenen Automatisierungsfehlern) ein höherer Workload (i.S. einer Überbeanspruchung) vorherrschen. Dies konnte nicht bestätigt werden. Als Erklärung muss das Signal-Rausch-Verhältnis, also die Anzahl der Automatisierungsfehler relativ zur Fahrdauer ohne Fehlerpräsentation, hinterfragt werden: Selbst in der Bedingung „niedrige Funktionszuverlässigkeit“ traten verhältnismäßig wenige „Signale“ in einem großen Abstand von ca. acht Minuten auf. Diese niedrige Signalrate ist untypisch für klassische Vigilanzparadigmen. Ein einziges Signal innerhalb einer Stunde zu präsentieren, ist noch weniger mit typischen Vigilanzaufgaben vereinbar. Um eine hohe externe Validität (i.S. einer realitäts- bzw. seriennäheren Funktionsauslegung) zu erreichen, wurde in der Bedingung „niedrige Funktionszuverlässigkeit“ bewusst diese niedrige Signalrate gewählt. Die Bedingung „hohe Funktionszuverlässigkeit“ sollte darüber hinaus in der Lage sein, eine kontinuierliche Ermüdung ohne etwaige Störungen durch Fehler sichtbar zu machen. Diese niedrigen Signallraten können im Experiment jedoch bedingt haben, dass faktisch kein Beanspruchungsunterschied für die Probanden bestand – deshalb liegt der gemessene Workload zwischen beiden Bedingungen auch auf einem vergleichbaren Niveau. Ein anderer Erklärungsansatz ergibt sich aus der Salienz der Reize, bzw. der Signale. Auf das Experiment angewandt entspricht die Salienz u.a. den Längs- und Querbeseleunigungen, die der Proband während der Präsentation der Automatisierungsfehler wahrnehmen konnte. Bei hoher Salienz (vgl. Müdigkeitsmodell) sind die Fehler leichter wahrzunehmen – es bedarf keiner großen Anstrengung, diese nicht zu verpassen. Die konkret vorherrschenden Beseleunigungen wurden hier nicht gemessen. Trotzdem wird angenommen, dass speziell die Fehler der Längsführung einfacher wahrzunehmen waren als die Querablagen. Letztere wurden vom Wizard-Fahrer schleichend eingeleitet. Das Aufspüren der Fehler in der Längsführung gestaltete sich unter Umständen leichter, weil dem Fahrer eine genaue Statusanzeige (nämlich der Tachometer) für etwaige Änderungen der Fahrgeschwindigkeit zusätzlich zur vestibulären Wahrnehmung zur Verfügung stand. Mit anderen Worten war es für die Probanden einfacher, eine Überschreitung der 80 km/h-Markierung der Tachonadel zu bemerken als genau den Moment der Überschreitung der Fahrstreifenmarkierung zu bestätigen. Womöglich ist also davon auszugehen, dass die Salienz der Automatisierungsfehler in Summe zu hoch war, als dass sich eine Überbeanspruchung

abzeichnen konnte. Generell passt diese Argumentation gut zu dem Erklärungsversuch von Stapel et al. (2017), die eine Level 2-Fahrt mit einem Tesla Modell S durchführten. Demnach war die geforderte Überwachungstätigkeit sehr wohl höchst beanspruchend, jedoch bewirkte ein hoher Monotoniegrad der Verkehrsumgebung, dass die Versuchsteilnehmer mehr Ressourcen für die Vigilanzaufgabe aufbringen konnten – so war die Überbeanspruchung nicht mehr messbar.

Parallel zum Workload sollten seltene Automatisierungsfehler geringer Salienz mit einem hohen Stresserleben einhergehen. Davon kann deskriptiv in der Messung nach der Versuchsfahrt ausgegangen werden, wenn Probanden trotz randomisierter Zuweisung zur Bedingung „niedrige Zuverlässigkeit“ nicht a priori ein signifikant höheres Stresserleben besessen hätten. Zusätzlich war die Stressmessung durch den Persönlichkeitstrait *Neurotizismus* konfundiert, da Probanden mit dieser Ausprägung ein signifikant höheres Stresserleben vermerkten. In der Summe kann daher keine fundierte Aussage zum Zusammenhang zwischen der Anzahl der Automatisierungsfehler und dem Stresserleben getätigt werden.

Obwohl in der Bedingung „niedrige Funktionszuverlässigkeit“ die Versuchsteilnehmer alle acht Minuten auf einen Fehler in der Längsführung oder auf eine Querlablage reagieren mussten, ist deskriptiv eine Tendenz zu beobachten, wonach die Teilnehmer die Nutzung einer solch „schlechteren“ Funktion leicht bevorzugen. Eine zufallskritische Absicherung dieses Haupteffekts war schlichtweg nicht möglich – aus diesem Grund sollte die Interpretation mit größter Vorsicht erfolgen. Es könnte lediglich geschlussfolgert werden, dass die Probanden eine regelmäßige Durchbrechung der Monotonie nicht ablehnen. Allerdings besaßen die Teilnehmer keine direkte Vergleichsmöglichkeit zur „hohen Funktionszuverlässigkeit“ aufgrund des Zwischengruppendesigns. Auch der gemessene Komfort, das Vertrauen in die Fahrfunktion, sowie das Befinden nach der Fahrt (i.S. einer Ausbildung von Kinetose) unterschied sich nicht zwischen den beiden erlebten Zuverlässigkeiten.

Nun sollen die Unterschiede zwischen den Probanden betrachtet werden, die anhand ihrer EEG-Alpha-Spindelrate als „ermüdet“ oder „nicht ermüdet“ klassifiziert wurden. Hier fällt zunächst der signifikant geringere globale, wie auch der signifikant geringere kognitive Workload der „ermüdeten“ Teilstichprobe auf. Damit besteht ein weiterer Hinweis, dass entsprechend des angepassten Müdigkeitsmodells eine Unterbeanspruchung in Zuge der fortwährenden Überwachung der Level 2-Funktion stattgefunden hat. Der hypothesenkonforme Trend im gemessenen Stresserleben stützt diese Vermutung, da „ermüdete“ Probanden ein geringfügig reduziertes Stresserleben vermerkten (auch hier sollte die Konfundierung durch Neurotizismus beachtet

werden). Leider konnte nicht nachgewiesen werden, dass Probanden mit höherem Funktionsvertrauen aufgrund der sich ergebenden Langeweile eine Unterbeanspruchung und damit eine stärkere Ermüdung zeigen. Darüber hinaus konnte zwar gezeigt werden, dass die Nutzungsbereitschaft mit dem Erleben der Level 2-Funktion signifikant ansteigt, jedoch empfanden objektiv „ermüdet“ klassifizierte Probanden keine signifikante Aversion, aufgrund derer sie die Funktion nicht nutzen würden. Ein Trend in die entsprechende Richtung, wonach die Nutzungsbereitschaft mit höherer Müdigkeit fällt, ist marginal erkennbar. Für den empfundenen Komfort fallen analog zur Variable „Funktionszuverlässigkeit“ keine Unterschiede auf.

In der Summe konnten im ersten Experiment keine Merkmale einer Überbeanspruchung durch eine Level 2-Nutzung nachgewiesen werden. Das heißt nicht zwangsläufig, dass diese vigilanzaufgabentypischen Merkmale im Anwendungskontext gar nicht existieren können. Ein entscheidender Faktor könnte einerseits die Anzahl von Automatisierungsfehlern relativ zur Fahrdauer darstellen, andererseits die Salienz der Fehler. So mag es der experimentellen Konzeption geschuldet sein – die es wiederum ermöglichte, andere Effekte sichtbar zu machen –, dass Merkmale von Überbeanspruchung nur eine geringe Chance besaßen, überhaupt in Erscheinung zu treten. Hier wäre eine mit Abwandlungen konzipierte Studie wünschenswert – es ist also weiterer Forschungsbedarf angezeigt. Umgekehrt gibt es verschiedene Hinweise, die über eine Unterbeanspruchung zu einer passiven, aufgabenbezogenen Müdigkeit geführt haben. Allen voran ist hier der signifikant geringere kognitive Workload bei „ermüdet“ klassifizierten Versuchsteilnehmern zu nennen. Dieser lässt entsprechend des angepassten Müdigkeitsmodells auf ein zu niedriges Arousal-Level schließen. Hier ist demnach davon auszugehen, dass die vorherrschende Monotonie der Teststrecke im Zusammenwirken mit der gestellten Aufgabe (Überwachung eines Level 2) die Unterbeanspruchung bewirkte. Die Folge war eine höhere passive aufgabenbezogene Müdigkeit, die zum Vigilanzdekrement führte. Über das gemessene Funktionsvertrauen konnte jedoch die modelltheoretisch hypothetisierte Langeweile im Kausalzusammenhang mit einer höheren Müdigkeit nicht abgebildet werden.

Betrachtet man die im Modell verankerte schlafbezogene Müdigkeit genauer, so konnten jedoch verschiedene Ursachen hierfür aufgespürt werden. Insbesondere sind hier der zuvor diskutierte Einfluss des Koffeins auf die Müdigkeit als auch die Schlafqualität zu nennen – wenn man diese über Umwege über die Anzahl der Schlafstunden argumentiert. Bezogen auf den Zirkadianen Rhythmus und den Chronotyp der einzelnen Probanden lässt das erste Experiment nur Vermutungen zu. Abschließend sollen die Stärken und Schwächen des verwendeten Untersuchungsansatzes diskutiert werden, bevor die Fragestellungen des ersten Experiments zusammenfassend beantwortet werden.

Untersuchungsansatz und Versuchskonzeption

Grundlegend ist festzustellen, dass der Einsatz des Wizard-of-Oz-Fahrzeugs in vorliegender Untersuchung mit nahezu perfekter Täuschung realisiert werden konnte, da nur ein einziger Proband Mutmaßungen über einen zweiten Fahrer äußerte. Die Immersionsgüte im Fahrversuch kann entsprechend hoch beziffert werden. Entsprechend der Ergebnisse wurde die Ermüdungsstärke nicht durch die Steuerungsgüte des Wizard-Fahrers beeinflusst. Abstriche bei dieser Interpretation müssen in der Form gemacht werden, als dass die SDLP über die gesamte Runde erhoben wurde, also auch in den Steilkurven, in denen die Messung womöglich verzerrt war. Insgesamt konnten aber die wesentlichen methodischen Herausforderungen nach Müller et al. (2019) zur Erfüllung der einschlägigen Testgütekriterien erfolgreich bewältigt werden.

Die Kombination aus Wizard-of-Oz-Fahrzeug und Teststrecke überzeugte dadurch, dass ein experimenteller Versuchsplan mit vergleichbarer Reliabilität wie in einem Fahrsimulator umgesetzt werden konnte. Dabei profitierte die Versuchvalidität von der Fahrt in einem realen Fahrzeug. Die Kontrollierbarkeit war hoch, ebenso konnte eine hohe Monotonie durch die Teststreckencharakteristika sichergestellt werden. Obwohl somit der Spagat zwischen hoher interner und externer Validität geglückt war, soll hier im Speziellen die Übertragbarkeit auf den realen Straßenverkehr diskutiert werden. Auf den ersten Blick könnte kritisiert werden, dass bei Level 2-Nutzung auf einer Bundesautobahn aufgrund deutlich geringerer Monotonie, zumal weitere Verkehrsteilnehmer anwesend sind, eine langsamere und/oder gedämpfte Ermüdung zu beobachten sein könnte – sodass die externe Validität des ersten Experiments zu niedrig war. Auf den zweiten Blick muss hier in entgegengesetzter Richtung argumentiert werden: Mit hoher Wahrscheinlichkeit werden Fahrer eine Level 2-Funktion genau dann nutzen, wenn die Fahrsituation aufgrund ihrer Komplexität eine Aktivierung lohnenswert erscheinen lässt, ohne dass zu viele Rückübernahmen zur manuellen Fahrzeugsteuerung drohen. Unter dieser Prämisse besteht die Vermutung, dass Fahrer sich in eben diesen monotonen, gar langweiligen Fahrsituationen eine Unterstützung durch eine automatisierte Fahrfunktion herbeiwünschen. Handelt es sich hierbei jedoch um eine Level 2-Funktion, wie im vorliegenden Experiment, ist mit den aufgezeigten Leistungseinbußen über die Fahrdauer zu rechnen. Insofern konnte das erste Experiment – mit hohem Aufwand in der Durchführung – wertvolle, weil objektive Einblicke in den Fahrerzustand bei Verwendung eines Level 2 gewähren, die mitunter kritisch in Hinblick auf die Verkehrssicherheit zu beurteilen sind.

Wie in so vielen Studien, die sich in den letzten Jahren mit der Fahrermüdigkeit beim automatisierten Fahren beschäftigt hatten, musste sich auch im vorliegenden Experiment mit erheblichen intra- und interindividuellen Unterschieden im Ermüdungsverhalten auseinandergesetzt werden. Um diese kontrollieren zu können, wäre a) eine größere Stichprobe oder b) eine differenzierte Probandenauswahl vonnöten gewesen. Aufgrund ökonomischer Restriktionen war Variante a) unrealistisch, Variante b) hätte im Umkehrschluss die Repräsentativität der Untersuchung geschmälert. Dennoch konnte hier „aus der Not eine Tugend“ gemacht werden, indem bestimmte individuelle Einflussfaktoren als Validitätsaußenkriterien herangezogen wurden. Die Müdigkeit selbst konnte „nicht invasiv“, also ohne Störung durch eine Abfrage, mittels EEG und Videobildern objektiv erfasst werden.

Ein besonderes Alleinstellungsmerkmal neben der Kombination aus EEG-basierter Müdigkeitsdetektion in einen Wizard-of-Oz-Fahrzeug unter kontrollierten Teststreckenbedingungen, bestand in der Darbietungsweise der Automatisierungsfehler. In Studien von Körber et al. (2015) oder Stapel et al. (2017) wurde eine sekundäre Vigilanzaufgabe (auditives Oddball-Paradigma bzw. DRT) verwendet, um hierüber Müdigkeitsableitungen zu tätigen. Die Nachteile dieser nicht störungsfreien Messungen wurden in Abschnitt 2.4.5.1 dargelegt. Die automatisierte Fahrt durch eine Transition vollständig zu unterbrechen, wie es Schömig et al. (2015) umsetzten, ist noch restriktiver. Im vorliegenden Experiment konnte hingegen von einer Integration von „Ereignissen“, d.h. Automatisierungsfehlern mit hoher externer Validität in den logischen Kontext einer Level 2-Überwachung profitiert werden. Leider detektierte die Sensor-kamera des Versuchsfahrzeugs die Fahrstreifenüberschreitungen in den Steilkurven nur unzuverlässig, sodass keine Reaktionszeitentwicklung im Hinblick auf die Automatisierungsfehler über die Fahrdauer ausgewertet werden konnte. Auch bestand zwischen Fehlern in Längsführung und Querablagen eine unterschiedliche Salienz. Dennoch kann – bei entsprechender Berücksichtigung der Problemquellen – für zukünftige Forschungsvorhaben ein solcher Ansatz empfohlen werden.

Im anschließenden Experiment sollen Maßnahmen evaluiert werden, um den Fahrerzustand auf einem verkehrssicheren Niveau zu halten. Konkret wird das Aktivierungspotenzials eines Level 3 unter Bearbeitung motivierender fahrfremder Tätigkeiten untersucht und mit dem bereits bekannten Level 2 verglichen werden. Ein besonderer Fokus liegt dabei auf den Auswirkungen von Übernahmeaufforderungen (Transitionen) zum manuellen Fahren auf den Fahrerzustand – wobei auf Grundlage der bisherigen Ergebnisse und der vorangegangenen Diskussion von einem hohen Aktivierungspotential dieser Transitionen – zumindest nach Level 2 – ausgegangen wird.

4 Experiment 2

4.1 Methode

Das zweite Experiment zielte auf die Erfassung des Fahrerzustands, konkret der Müdigkeit, während einer automatisierten Realfahrt (sowohl SAE Level 2 als auch Level 3) ab. Neben dem Wizard-of-Oz-Forschungsfahrzeug auf der Teststrecke kam analog zum ersten Experiment auch die Alpha-Spindeldetektion mittels EEG als primärer Müdigkeitsindikator zum Einsatz. Das Versuchsdesign sollte Vergleiche zwischen einer Level 2-typischen Überwachungsaufgabe und einer Level 3-typischen Beschäftigung mit fahrfremden Tätigkeiten (FFT) ermöglichen. Da große Teile der verwendeten Untersuchungsmethodik kongruent zum ersten Experiment waren, wird sich in den folgenden Abschnitten auf Unterschiede bzw. Anpassungen der experimentellen Methodik beschränkt.

4.1.1 Versuchsplan

Im zweiten Fahrversuch sollte das Ermüdungsverhalten in Abhängigkeit von der automatisierten Fahrdauer (erste unabhängig Variable; UV 1) und des Automatisierungslevels (UV 2) untersucht werden. Im Gegensatz zum ersten Experiment wurden keine Aggregationen der minütig aufgezeichneten Alpha-Spindeln vorgenommen. Insofern umfasst die Anzahl der Faktorstufen von UV 1 die Anzahl der messbaren Fahrminuten. Eine ununterbrochene automatisierte Fahrperiode besaß eine Distanz von 15 Runden auf der Teststrecke, sodass eine Fahrdauer von ca. 30 Minuten anvisiert wurde. Die Mindestfahrdauer, die bei jedem Probanden erreicht werden sollte und für den Versuchsplan zugrunde gelegt wurde, betrug 28 Minuten. Unter Verwendung eines um 300 s gleitenden Mittelwerts für die Alpha-Spindeldetektion, lag der erste „reine“ Wert automatisierter Fahrt bei Minute sechs vor. Daher konnte UV 1 maximal 23 Faktorstufen annehmen.

Alle Probanden sollten mindestens 28 Minuten in einem Level 2 und mindestens 28 Minuten in einem Level 3 fahren. Nach jeder automatisierten Fahrperiode erfolgte eine Übernahmeauforderung (Transition) zum manuellen Fahren, das jeweils ca. 2.5 Minuten andauerte (vgl. Abbildung 49).

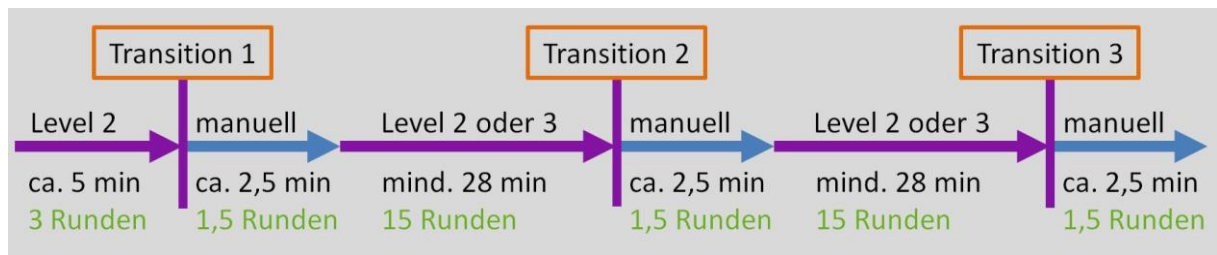


Abbildung 49: Abfolge der Versuchsfahrt in Experiment 2.

Diesen zwei 28-minütigen Fahrten war eine ca. vierminütige Level 2-Fahrt vorgelagert, nach der die erste Übernahmeaufforderung (Transition 1) erfolgte. Dies hatte zum Zweck, die konfundierende Erstreaktion der Probanden auf eine Übernahmeaufforderung, die sich typischerweise von nachfolgenden Reaktionen unterscheidet (vgl. Feldhütter et al., 2017), aus der eigentlichen Analyse auszuschließen. Daher waren im Versuchsplan erst Transitionen 2 und 3 von Analyseinteresse, die sich nach der Level 2- oder Level 3-Fahrt anschlossen. Die Hälfte der Probanden war zufällig der Reihenfolge A (erst Level 2, dann Level 3) und die andere Hälfte zufällig der Reihenfolge B (erst Level 3, dann Level 2) zugewiesen. Folglich wurde UV 2 messwiederholt mit zwei Faktorstufen (Level 2 und Level 3) erhoben. Dem zweiten Experiment lag demnach ein 23×2 faktorielles Design mit Messwiederholung auf beiden Faktoren zugrunde. Eine Übersicht aller erhobenen abhängigen Variablen findet sich in Tabelle 16.

Tabelle 16: Subjektive abhängige Variablen im Experiment 2 inklusive Messzeitpunkt und Bezug zum jeweiligen Automatisierungslevel.

subjektive abhängige Variable	vor dem Fahrversuch		nach dem Fahrversuch	
	generell	Level 2 / 3	generell	Level 2 / 3
Müdigkeitseinschätzung	×		×	
Nutzungsbereitschaft		×		×
Vertrauen in die Automatisierung		×		×
Stressempfinden	×			×
Workload, mit / ohne Transition				×
Schlafdauer vor der Versuchsfahrt			×	
Koffeinkonsum am Morgen			×	

4.1.2 Stichprobe

Insgesamt nahmen 19 Probanden an dem Versuch teil. Da bei drei Probanden verrauschte EEG-Daten aufgezeichnet wurden, ein Proband maßgeblich von den Instruktionen abwich, und es bei einem Probanden während der Versuchsfahrt zu einer Kollision mit einer Taube auf der Windschutzscheibe kam, wurden diese vollständig aus der Auswertung ausgeschlossen. Die Nettostichprobengröße betrug daher $N = 14$. Hiervon waren $n = 9$ der Reihenfolge A und $n = 5$ der Reihenfolge B zugeordnet. Die Zuweisung zur jeweiligen Reihenfolge erfolgte so, dass eine annähernde Vergleichbarkeit der Reihenfolgen hinsichtlich Alter und Geschlecht der Probanden gewährleistet war. Reihenfolge A waren somit vier weibliche und fünf männliche Probanden im Alter zwischen 18 und 61 Jahren ($M = 40$, $SE = 5$) zugewiesen. Sie hatten im Mittel seit 23 Jahren ($SE = 5$) eine Pkw-Fahrerlaubnis und fuhren durchschnittlich ca. 30100 km ($SE = 10600$) im Jahr. In Reihenfolge B wurden zwei weibliche und drei männliche Probanden im Alter zwischen 19 und 55 Jahren ($M = 42$, $SE = 6$) untersucht. Sie hatten im Mittel seit 25 Jahren ($SE = 6$) eine Pkw-Fahrerlaubnis und fuhren durchschnittlich ca. 15700 km ($SE = 3200$) im Jahr.

4.1.3 Probandenakquise

Analog zum ersten Experiment wurden alle Probanden per Zeitungsannonce auf die Studie aufmerksam gemacht. Das Inserat wurde in der Samstagsausgabe der Aachener Zeitung (nun erweitert um die Bezirksausgabe „Aachen Stadt & Land“) mit folgendem Text geschaltet: „Studie „automatisiertes Fahren“: Die Bundesanstalt für Straßenwesen (BASt) sucht Personen im Alter zwischen 18 und 65 Jahren, die Interesse haben, einmalig (Termine Mo. bis Fr.) an einer EEG-Studie zum automatisierten Fahren teilzunehmen. Personen, die bereits letztes Jahr an der Studie teilgenommen haben, können leider nicht berücksichtigt werden. Ort: Aldenhoven Testing Center, Zeitaufwand: ca. 4 Std., Vergütung: 80 EUR. Bei Interesse melden Sie sich bitte unter [...]“. Hieraufhin meldeten sich 46 Personen, wobei die Hälfte der Interessenten zwischen 50 und 65 Jahren alt waren. Die Vertragsbedingungen der Unfallversicherung sahen weiterhin eine Altersbegrenzung von 65 Jahren vor. Es galten die gleichen Teilnahmevoraussetzungen wie für Experiment 1. Als Cover-Story wurde den Probanden erläutert, dass sie – allgemein gehalten – automatisiertes Fahren in einem Forschungsfahrzeug der BASt erleben würden, um Rückmeldung über die Funktionsweise des Systems zu geben, damit eine Optimierung der Automatisierungsfunktion stattfinden kann.

Es wurde das Ziel verfolgt, den Probanden einen Versuchstermin zuzuweisen, zu dem eine erhöhte Wahrscheinlichkeit zur Ermüdung bestünde. Daher wurde mittels vier Items des D-MEQ (Griefahn, Künemund, Bröde & Mehnert, 2001) als validierte deutsche Übersetzung des „Morningness-Eveningness-Questionnaires“ (MEQ) nach Horne und Östberg (1976) bei der Terminvergabe am Telefon kurz abgeprüft, ob der Versuchsteilnehmer eher einem Morgen- oder eher einem Abendtyp entsprechend seines Chronotyps zuzuordnen ist. Einem Morgentyp wurden in der Konsequenz ausschließlich Termine am späten Nachmittag (Beginn der Versuchsfahrt um 17 Uhr, einem Abendtyp ausschließlich Termine am Vormittag (Beginn der Versuchsfahrt um 11 Uhr) angeboten. Ein noch späterer Beginn für den Morgentyp bzw. ein noch früherer Beginn für den Abendtyp wäre wünschenswert gewesen (vgl. Abschnitt 2.3.3.2), war jedoch aufgrund der exklusiven Nutzung der Ovalbahn aus organisatorischen Gründen seitens des Teststreckenbetreibers nicht dauerhaft zu realisieren.

4.1.4 Versuchsmaterial

Sowohl die Teststrecke als auch das Messsystem zur Ableitung des EEG wurden kongruent zum ersten Experiment eingesetzt. Daher wird an dieser Stelle auf die Abschnitte 3.1.4.1 und 3.1.4.3 verwiesen.

4.1.4.1 Versuchsfahrzeug

Für das zweite Experiment stand dasselbe Wizard-of-Oz-Fahrzeug zur Verfügung, das auch im ersten Experiment zum Einsatz kam. Allerdings wurden umfangreiche Umbauten und Erweiterungen durchgeführt. Für die Erklärungen empfiehlt es sich Abbildung 50 heranzuziehen.



Abbildung 50: Fondbereich des „Wizard-of-Oz-Fahrzeugs“ aus Sicht des Wizard-Fahrers nach Erweiterungen für das zweite Experiment.

Oberhalb der Bedienkonsole im Fond war nun ein Touchscreen (kapazitives Display) angebracht, über den der Versuchsleiter eine Geschwindigkeitsregelanlage bedienen konnte. Mit deren Hilfe konnte eine konstante Fahrgeschwindigkeit von 70 km/h in den Automatisierungsphasen gewährleistet werden. Links oberhalb der zentralen Bedienkonsole wurde ein Tablett-PC befestigt. Über große Berührungsflächen konnten hier Zustände der Fahrzeugautomatisierung an ein Display im Kombiinstrument vor dem Probanden gesendet werden. Außerdem befand sich links vom Versuchsleiter ein zusätzlicher Messrechner, der eine Aufzeichnung des jeweils vorherrschenden Automatisierungszustands sicherstellte und über CAN-Bus an den EEG-Messrechner auf dem Beifahrerplatz übertrug.

In der vorderen Fahrerkabine wurden die Videokameras an denselben Positionen befestigt, wie in Experiment 1. Somit erlaubte die Ausrichtung der Kamera auf Oberschenkelhöhe des Probanden, die Beschäftigung mit fahrfremden Tätigkeiten aufzuzeichnen. Auf dem EEG-Messlaptop wurde ein grünes Ablagefach für die vom Probanden mitgebrachten fahrfremden Tätigkeiten angebracht (vgl. Abbildung 51). In diesem lagen zusätzlich mehrere Zeitschriften für den Fall bereit, dass ein Proband ohne Tätigkeit zum Fahrversuch erschien. So war sichergestellt, dass er sich in jedem Fall mit fahrfremden Tätigkeiten während der Level 3-Periode beschäftigen konnte.



Abbildung 51: Blick auf den vorderen Probandenplatz. Links: Mensch-Maschine-Interaktionskonzept für den Probanden mit der automatisierten Fahrfunktion. Rechts: Ablagefach für fahrfremde Tätigkeiten oberhalb des EEG-Messrechners.

Das Kombiinstrument wurde um eine Anzeige des aktuell vorherrschenden Automatisierungszustands erweitert (vgl. Abbildung 51). Als Anzeigedisplay diente ein Smartphone mit Android-Betriebssystem, das zwischen Tachometer und Drehzahlmesser Platz fand – wodurch lediglich die Anzeige im Kombiinstrument (der Bordcomputer) überdeckt wurde. Eine vollständige Übersicht aller Automatisierungszustände findet sich in Abbildung 52.

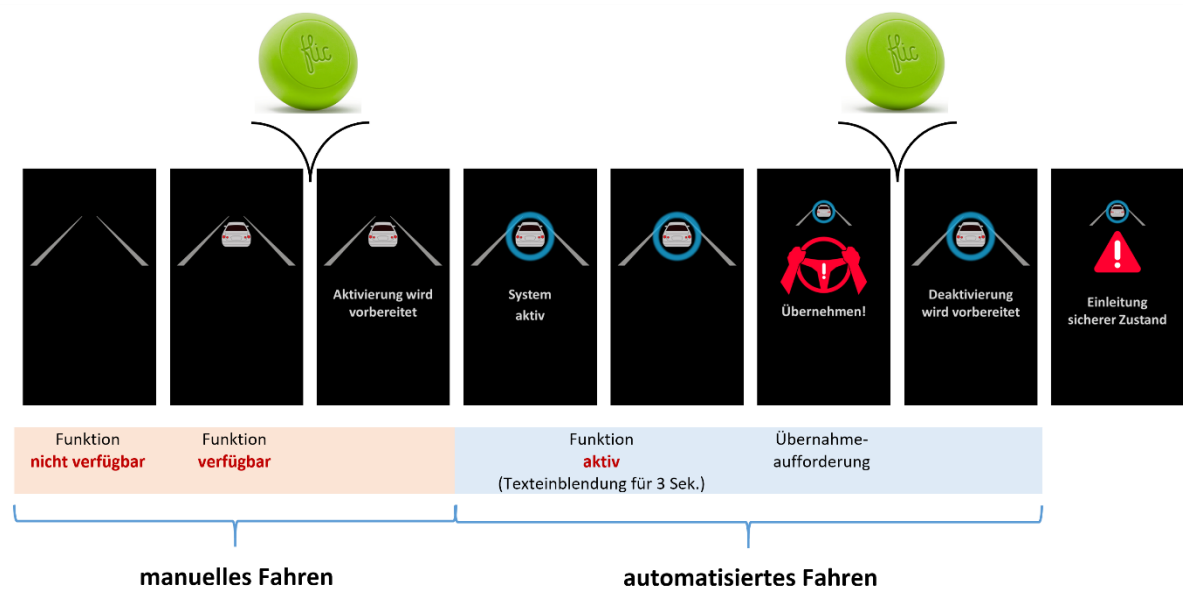


Abbildung 52: Darstellbare Automatisierungszustände im Kombiinstrument des „Wizard-of-Oz-Fahrzeugs“ (aus Klamroth et al., 2019).

Auf dem Lenkrad wurde unterhalb der rechten Daumenablage ein grüner Taster („Flic Smart Button“) mit einem Durchmesser von 28 mm (Druckbereich 20 mm Ø) angebracht, der über Bluetooth mit dem Android-Smartphone verbunden war. Der Proband konnte mit einem Druck auf den Taster die Automatisierungsfunktion aktivieren und deaktivieren. Die Wechsel zwischen der Fahrzeugsteuerung vom Probandenplatz und vom zweiten Fahrerplatz im Fond wurden wie folgt gestaltet: Der Versuchsleiter bot dem Probanden die Automatisierungsfunktion zum definierten Zeitpunkt (vgl. Abbildung 52) an, sodass im Kombiinstrument das Icon „Funktion verfügbar“ erschien. Entsprechend der Instruktion aktivierte der Proband schnellstmöglich die Automatisierungsfunktion mit einem Druck auf den grünen Taster. Der Versuchsleiter erhielt daraufhin einen Hinweis, dass der Proband zum automatisierten Fahren bereit ist. Daraufhin aktivierte der Versuchsleiter die Geschwindigkeitsregelanlage und schloss die Magnetkupplung der Lenkung. Während dieser Tätigkeiten wurde dem Probanden angezeigt, dass die Automatisierung vorbereitet wird. Mit Schließen der Magnetkupplung lag die Fahrzeugsteuerung vollständig in der Hand des Versuchsleiters, dementsprechend wurde auf dem Probandendisplay angezeigt, dass die Automatisierung nun aktiv ist. Sobald das Ende der automatisierten Fahrperiode erreicht war, gab der Versuchsleiter dem Probanden die Aufforderung zum manuellen Fahren. Hierfür berührte er den entsprechenden Befehl auf dem Tablett-PC und der Proband erhielt eine Übernahmeaufforderung auf dem Display gekoppelt mit einem Ton von 480 ms Dauer, dessen Frequenzpeak bei 439 Hz lag.

Entsprechend der Instruktion übernahm der Proband die manuelle Fahrzeugsteuerung – sobald er hierzu subjektiv in der Lage war – indem er wiederum den grünen Taster betätigte. Für die Erhebung der Reaktionszeit auf eine Übernahmeaufforderung hin, wurde die Zeitdifferenz zwischen der Aufforderung durch den Versuchsleiter und der probandenseitigen Bestätigung (Zeitpunkt des Tastendrucks) gemessen. Nach erfolgreichem Tastendruck zur Deaktivierung der Automatisierungsfunktion erhielt der Versuchsleiter auf dem Tablett-PC den entsprechenden Hinweis, dass der Proband seine bevorstehende manuelle Fahrzeugsteuerung bestätigt hat. Der Versuchsleiter deaktivierte die Geschwindigkeitsregelanlage und öffnete wieder die Magnetkupplung der Lenkung. Während dieses Vorgangs sah der Proband auf dem Display, dass die Deaktivierung der Automatisierung vorbereitet wird. Mit dem Öffnen der Magnetkupplung lag die vollständige Fahrzeugsteuerung wieder beim Probanden, entsprechend wurde dem Probandendisplay der manuelle Fahrbetrieb angezeigt. Ausschließlich für den Fall, dass der Proband trotz Übernahmeaufforderung nach 20 Sekunden nicht reagiert hätte, war es dem Versuchsleiter möglich, mittels Einblendung „Einleitung sicherer Zustand“, den Versuch abzubrechen, ohne das „Wizard-of-Oz-Fahrzeugkonzept“ gegenüber dem Probanden offen zu legen.

4.1.4.2 Fragebögen

Auch im zweiten Experiment sollten sowohl subjektive Daten der Versuchsteilnehmer vor als auch nach der Versuchsfahrt erhoben werden. Die vorab abgefragten deskriptiven Daten, die die Stichprobenkonstellation beschreiben, wie auch die Erfahrung mit der Verwendung von Fahrerassistenzsystemen wurde analog zum ersten Experiment erfasst. Die Fragebögen finden sich in Anhang C und D und können zu den Erläuterungen hinzugezogen werden. Bevor die Probanden Erfahrung mit einer kontinuierlichen Fahrzeugautomatisierung Level 2 und Level 3 sammeln konnten, wurden sie vorab mittels SOMS (Reinprecht, 2013) nach ihrer subjektiv empfundenen Müdigkeit gefragt. Anschließend erhielten die Versuchsteilnehmer sowohl eine Kurzinformation zum automatisierten Fahren in Level 2 als auch Level 3. Kernelement dieser Information war die Rollenbeschreibung von Funktion und Fahrer hinsichtlich ihrer Leistungsfähigkeit und Aufgaben. Ausgehend hiervon sollten die Probanden abschätzen, ob sie eine solche Fahrfunktion in ihrem Alltag verwenden würden und welches Vertrauen sie dem jeweiligen Level automatisierten Fahrens entgegenbringen würden. Bewusst wurde auch im zweiten Fahrversuch darauf verzichtet, die Einteilung in verschiedene Level automatisierten Fahrens gegenüber den Probanden einzuführen. Stattdessen wurde das Augenmerk auf das Rollenverständnis innerhalb der Mensch-Maschine-Interaktion seitens des Probanden gelegt, indem der Level 2 als „Automatisiertes Fahren mit Überwachung“ und der Level 3 als „Automatisiertes Fahren mit fahrfremden Tätigkeiten“ bezeichnet wurde. Die Vertrauensabfrage wurde – entsprechend

Experiment 1 – mittels „Automation Trust Scale“ nach Jian et al. (2000) durchgeführt. Abschließend sollte das Stress-Niveau der Probanden vor dem Versuch wiederum mit Hilfe des SSSQ (Helton, 2004) – mit der unter Abschnitt 3.1.4.4 beschriebenen Anpassung – und die Big-Five Persönlichkeitsstruktur über den BFI-10 nach Rammstedt und John (2007) erhoben werden.

Unmittelbar im Anschluss an den zweiten Fahrversuch schloss sich eine umfangreiche Nachbefragung der Probanden an (Fragebogenheft Teil 2, Anhang D), die mit subjektiver Einschätzung der Müdigkeit anhand der SOMS einleitete. Hierauf folgten Fragen zur Nutzungsbereitschaft, NASA-TLX, „Automation Trust Scale“ und SSSQ (mit den zuvor genannten Anpassungen), wobei sich zunächst die gesamte Befragung auf das Fahren mit Level 2 und anschließend auf das Fahren mit Level 3 bezog. Sowohl nach der 28-minütigen Level 2-Fahrt als auch nach der 28-minütigen Level 3-Fahrt fand eine Übernahmeaufforderung zum manuellen Fahren statt (vgl. Abschnitt 4.1.1). Da nicht nur der Workload der Probanden während der automatisierten Fahrt, sondern auch während der Übernahme-situation interessierte, wurde hinsichtlich des NASA-TLX eine weitere Unterteilung vorgenommen: Die Versuchsteilnehmer sollten zunächst die automatisierte Fahrperiode *vor* Übernahme der Fahrzeugsteuerung bewerten, und anschließend *nur* die Übernahme-situation an sich beurteilen. Abschließend erfolgte kongruent zu Experiment 1 die Abfrage der Wizard-of-Oz-Immersionsgüte sowie die Angabe der geschlafenen Stunden und einer im Vergleich zum ersten Experiment differenzierteren Angabe des Koffeinkonsums vor dem Fahrversuch.

4.1.5 Instruktion

Der Proband wurde analog zu Experiment 1 schriftlich über seine Aufgabe während des Fahrversuchs instruiert. Eingangs wurde grob erläutert, dass es Zielsetzung der Studie sei, Auswirkungen von automatisiertem Fahren auf den Fahrer zu untersuchen und eine Funktionsbewertung vorzunehmen. Sowohl automatisierte als auch manuelle Fahrperioden wurden dem Probanden angekündigt. Allerdings wurden die Versuchsteilnehmer auch in diesem Versuch über die genaue Dauer der jeweiligen Fahrperioden im Unklaren gehalten.

Hinsichtlich der manuellen Fahrt wurde der Proband gebeten, konstant mit 70 km/h auf dem mittleren Fahrstreifen der Ovalbahn zu fahren. Kongruent zum ersten Experiment wurde die Funktionsweise der Automatisierung erklärt: Das Oval würde mit einer (leicht variablen) Geschwindigkeit zwischen 60 und 80 km/h auf dem mittleren Fahrstreifen durchfahren, wobei der Versuchsteilnehmer nie in die Automatisierung eingreifen sollte – es sollte folglich immer

„hands-off“ gefahren werden. Da das vermittelte Bild einer zweistreifigen Autobahn mit Seitenstreifen (wie in im ersten Fahrversuch verwendet) teilweise zu artifiziell wirkte, wurde im zweiten Experiment zugunsten der Informationssparsamkeit hierauf verzichtet, da der Proband eine Vielzahl weiterer Instruktionen erhielt, die das Anzeige-konzept im Fahrzeug betrafen. Die verschiedenen Zustände der (fiktiven) automatisierten Fahrfunktion, die auf dem Display im Kombiinstrument dargestellt werden konnten, wurden dem Probanden zur schriftlichen Instruktion beigelegt. Die Bebilderung (vgl. Abbildung 52) sollte vom Versuchsteilnehmer zum Nachvollziehen der Aktivierungs- und Deaktivierungsprozedur herangezogen werden.

Ohne auf eine konkrete Unterscheidung der Automatisierungslevel einzugehen, wurden in der schriftlichen Instruktion stattdessen die zwei verschiedenen Aufgaben des Probanden während des Versuchs kompakt dargestellt. Einerseits erfuhr der Versuchsteilnehmer, dass er bei Aufgabe 1 die automatisierte Fahrt permanent überwachen sollte. Sobald er Automatisierungsfehler bemerkte – die gleichermaßen wie in Experiment 1 definiert wurden, jedoch entsprechend des Versuchsplans nie in Erscheinung traten – sollte der Proband seinen Handtaster betätigen. Andererseits sollte er sich in Aufgabe 2 während der automatisierten Fahrperiode mit fahrfremden Tätigkeiten beschäftigen, die der Versuchsteilnehmer hierfür zum Versuch mitbrachte (und die damit vermeintlich motivierend sein sollten). Hier ist zu erwähnen, dass der Versuchsleiter mündlich darauf hinwies, möglichst permanent einer oder mehrerer Tätigkeiten im Wechsel nachzugehen. Hiermit sollte sichergestellt sein, dass die vorgesehene experimentelle Manipulation (Fahren in einem Level 3) tatsächlich auch umgesetzt werden kann. Die probandeneigenen fahrfremden Tätigkeiten sollten stets in der grünen Ablageschale auf dem Beifahrerplatz, oberhalb vom Messrechner (vgl. Abbildung 51, Abschnitt 4.1.4.1) aufbewahrt werden, und bei Aufgabe 2 zur Hand genommen werden. Bevor der Proband wieder manuell fuhr, sollten die Tätigkeiten wieder in diese Ablageschale zurückgelegt werden.

Auch für die manuelle Fahrt (Baselinefahrt) vor Beginn der automatisierten Fahrperioden, wurde der Proband analog zum ersten Fahrversuch instruiert. Hierfür war die Abfolge von Fahraufgaben auf der Mittelkonsole des Versuchsfahrzeugs befestigt (vgl. Anhang E).

4.1.6 Versuchsdurchführung

Das Experiment fand im Juli 2018 statt. Bei allen Versuchsfahrten herrschten trockene Bedingungen (kein Niederschlag, trockene Fahrbahn) mit konstant hoher Außentemperatur (zwischen 30 und 35°C) vor. Der Innenraum des Fahrzeugs wurde auf ca. 21°C klimatisiert.

Jeder Proband traf entsprechend seines angenommenen Chronotyps entweder gegen 09:30 Uhr oder 15:30 Uhr an der Teststrecke ein. Nach Begrüßung durch den Versuchsleiter schlossen sich die Vorbereitungen und Vorabbefragungen kongruent zu Experiment 1 an (vgl. Abschnitt 3.1.6). Die Zusammenstellung der Fragebögen für den zweiten Fahrversuch findet sich in Anhang C und D, die Instruktionen, die die Versuchsteilnehmer anschließend in schriftlicher Form erhielten, sind in Abschnitt 4.1.5 erläutert. Zu einem späteren Zeitpunkt (im Fahrzeug) wurden die Instruktionen dem Probanden erneut mündlich wiederholt. Insbesondere sollte für eine reibungslose Versuchsdurchführung sichergestellt sein, dass der Proband das Verfahren für die Aktivierung und Deaktivierung der Automatisierungsfunktion verstanden hatte. Daher wurde er gebeten, jeden einzelnen Zustand der in Abbildung 52 (vgl. Abschnitt 4.1.4.1) dargestellten Abfolge nochmals mit seinen eigenen Worten dem Versuchsleiter erklärend zu wiederholen. Machte der Proband hierbei Fehler, erklärte der Versuchsleiter einzelne Schritte erneut. So sollte bereits frühzeitig ein theoretisches Verständnis der Aktivierungs- und Deaktivierungsabfolge sichergestellt werden.

Der Proband nahm zügig im Versuchsfahrzeug Platz und wurde mit den spezifischen Sicherheitsmerkmalen vertraut gemacht. Im Fahrzeug bestand nun für den Probanden die Möglichkeit, sich jederzeit über die Gegensprechanlage mit dem Versuchsleiter im Fond zu verständigen. Die Verkabelung zwischen Messtechnik am Probanden und dem Messrechner auf dem Beifahrerplatz gestaltete sich wie im ersten Fahrversuch. Der Proband wurde gebeten, seine mitgebrachten fahrfremden Tätigkeiten in die Ablageschale zu legen. Bevor der Versuchsleiter den zweiten Fahrerplatz im verdunkelten Fond einnahm und den EEG-Messrechner startete, ging er den Ablauf des gesamten Versuchs mit dem Probanden zusammen durch und erkundigte sich nach Rückfragen.

Der Versuchsteilnehmer fuhr nach Anweisung des Versuchsleiters auf die Ovalbahn, absolvierte die EOG-Baselinemessung und die manuelle Fahrt über neun Runden anhand der Abfolge auf der Mittelkonsole (vgl. Anhang E) und hielt an der markierten Position auf der Teststrecke an (analog zu Experiment 1). Im Unterschied zum ersten Experiment sollte der Proband in den anschließenden zwei Runden Wechsel zwischen manuellem und automatisiertem Fahren und umgekehrt trainieren. Hierfür fuhr der Proband manuell an und beschleunigte das Versuchsfahrzeug auf 70 km/h. Sobald die Automatisierungsfunktion verfügbar war, aktivierte der Versuchsteilnehmer diese. Nach einer halben Runde Fahrt erhielt der Proband eine Übernahmeaufforderung zum manuellen Fahren. Dieser Ablauf wurde zweimal durchgespielt.

Nach der „Übungsfahrt“ brachte der Versuchsleiter das Fahrzeug wieder an der markierten Position zum Stillstand. Dem Probanden wurde kurz erläutert, dass nun der eigentliche Versuch beginne, der entsprechend der vorab gelesenen Instruktion verlaufen würde: Der Proband sollte das Fahrzeug auf 70 km/h beschleunigen. Bei Verfügbarkeit der Automatisierungsfunktion sollte diese schnellstmöglich aktiviert werden. Falls der Versuchsteilnehmer das Icon „Automatisierungsverfügbarkeit“ nicht wahrnahm, wies der Versuchsleiter den Probanden kurz über die Gegensprechanlage darauf hin, nun bitte die Funktion zu aktivieren. Unmittelbar nach Beginn der automatisierten Fahrt sagte der Versuchsleiter an, ob nun a) eine permanente Überwachung der Fahrgeschwindigkeit und Spurhaltung des Versuchsfahrzeugs notwendig war (entspricht einem Level 2) oder b) sich der Proband mit einer oder mehrerer fahrfremder Tätigkeiten beschäftigen sollte (entspricht einem Level 3). Die Dauer und die Reihenfolgen der automatisierten Fahrperioden sind im Versuchsplan verankert, siehe Abschnitt 4.1.1.

In diesem Experiment nutzte der Versuchsleiter eine Geschwindigkeitsregelanlage, um die Fahrgeschwindigkeit bei 70 km/h zu halten. Kongruent zum ersten Fahrversuch bekam der Versuchsleiter die aktuelle Runde angezeigt. Nach 36 Runden und nachdem der Proband die letzte Übernahmeaufforderung zum manuellen Fahren erhalten hatte, bat ihn der Versuchsleiter, das Fahrzeug wieder am markierten Standort auf der Ovalbahn zum Stillstand zu bringen. Unmittelbar darauf wurde die Datenaufzeichnung gestoppt und alle Messdaten in einer Gesamtdatei gespeichert. Der Proband wurde nun instruiert, die Teststrecke zu verlassen, nahm erneut im Büroraum Platz und bearbeitete sogleich den zweiten Teil des Fragebogens in Abwesenheit vom Versuchsleiter. Nach Ablegen des EEG wusch sich der Proband die Rückstände des Elektrolytgeles aus seinem Haar. Abschließend erläuterte der Versuchsleiter den fachlichen Hintergrund und klärte über die Täuschungen durch die Cover-Story und des Wizard-of-Oz-Fahrzeugkonzepts auf. Der Proband erhielt für die Versuchsdauer von ca. vier Stunden eine Aufwandsentschädigung in Höhe von 80 Euro und wurde verabschiedet.

4.1.7 Videoanalyse

Die Videoanalyse im zweiten Experiment verfolgte zwei Ziele. Einerseits sollte eine Replikation der Müdigkeitsbeurteilungen aus dem ersten Fahrversuch für die Level 2-Fahrt erfolgen. Andererseits interessierte auf deskriptiver Ebene für die Level 3-Fahrt, welche fahrfremden Tätigkeiten (FFT) Probanden wie lange ausführten, wie sie mit den Tätigkeiten bei Übernahmeaufforderung umgingen und wie häufig sie während der FFT-Beschäftigung Kontrollblicke auf die Umgebung richteten. Die Kontrollblicke sollten dabei als Indikator für Vertrauen in den Automatisierungsprozess fungieren (vgl. bspw. Hergeth et al., 2016).

Für die Müdigkeitsbeurteilung über Videorating mittels Mued 8.1 war die Definition einminütiger Videosequenzen der insgesamt mindestens 28-minütigen automatisierten Fahrperioden in Level 2 notwendig. Entsprechend des Versuchsplans war im Gegensatz zum ersten Experiment keine Einteilung in Zeitbereiche a priori vorgesehen (vgl. Abschnitt 4.1.1). Somit bestand die theoretische Möglichkeit, an beliebigen Zeitpunkten der automatisierten Fahrt das Rating vorzunehmen. Aus ökonomischen Gründen wurde sich für drei zeitliche Sequenzen entschieden. Wichtige Kriterien für die Auswahl dieser Zeitpunkte bestanden darin, dass die erste Sequenz nicht zu früh (d.h. nicht in den ersten fünf Minuten automatisierter Fahrt) liegen sollte, da hier noch keine „wahre“ EEG-Alpha-Spindelrate zum Vergleich vorlag. Daher wurde für die erste zu analysierende Sequenz der Bereich zwischen der sechsten und siebten automatisierten Fahrminute festgelegt. Außerdem sollte die letzte Sequenz nicht in die Phase der Übernahmeaufforderung zum manuellen Fahren fallen. Hier wurde entschieden, das Ende der letzten Videosequenz eine Minute vor der Übernahmeaufforderung probandenindividuell zu setzen, da keine exakt gleiche automatisierte Fahrdauer zwischen den Probanden vorlag (vgl. ebenfalls Abschnitt 4.1.1) – folglich startete die dritte einminütige Sequenz zwei Minuten vor Übernahmeaufforderung. Mit äquivalentem Abstand zwischen Messzeitpunkt eins und drei wurden Beginn und Ende der zweiten einminütigen Sequenz festgelegt. Zusammengefasst lag die erste einminütige Sequenz (t_1) zwischen Fahrminute sechs und sieben, die zweite (t_2) ungefähr um den Bereich der 17. Fahrminute und die dritte (t_3) um den Bereich der 27. Fahrminute. Der Abstand zwischen t_1 und t_2 , sowie zwischen t_2 und t_3 betrug demnach ca. zehn Minuten.

Alle Probandenvideos der Level 3-Fahrt wurden zunächst hinsichtlich der ausgeführten fahrfremden Tätigkeiten in Echtzeit angesehen. Dabei wurde vermerkt, welche Tätigkeit in welcher Zeitspanne ausgeführt wurde. Sobald die Übernahmeaufforderung zum manuellen Fahren erfolgte, wurde beobachtet, ob der Proband a) erst die Tätigkeit beiseitelegt, und dann die Übernahme bestätigt oder b) erst die Übernahme bestätigt, und dann die Tätigkeit beiseitelegt. Für die Kontrollblicke wurden insgesamt vier Zeitspannen definiert: 1-3 Sekunden, 4-6 Sekunden, 7-9 Sekunden und 10-15 Sekunden. Ein Blick auf die Anzeigen oder auf die Umgebung, der länger als 15 Sekunden andauerte, wurde als eine separate Tätigkeit („stetiger Blick“) klassifiziert.

4.2 Ergebnisse

4.2.1 Subjektive Urteile

Kein Proband äußerte die Vermutung, dass ein Wizard-Fahrer die Automatisierung des Versuchsfahrzeugs nachahmte. Die Schlafdauer der Stichprobe vor der Versuchsfahrt betrug im Mittel 6.61 Stunden ($SE = .25$). Gegeben, dass nach sieben Stunden keine Koffein-bedingten Effekte mehr nachweisbar sind (vgl. Weiß, 2007), zuvor jedoch schon, standen sechs Probanden während der Versuchsfahrt unter Koffeineinfluss, acht Probanden nicht. Die Vorerfahrung der Nettostichprobe ($N = 14$) mit SAE Level 1-Funktionen wird hier analog zu Experiment 1 angegeben. Demnach nutzten $n = 7$ Versuchsteilnehmer regelmäßig einen Tempomaten oder $n = 3$ einen Abstandsregeltempomaten. Bezogen auf Querregelung besaß ein Proband Erfahrung mit einer Parklenkfunktion, jedoch besaß niemand Erfahrung mit einer Spurhaltefunktion.

4.2.1.1 Müdigkeitseinschätzung

Abbildung 53 zeigt die mittlere subjektive Müdigkeit der Probanden vor und nach der Versuchsfahrt. Ein t -Test für abhängige Stichproben ergab, dass nach dem Fahrversuch eine signifikant höhere subjektive Müdigkeit als vor dem Fahrversuch vorlag ($t(11) = -4.42$, $p < .01$, $\eta_p^2 = .64$).

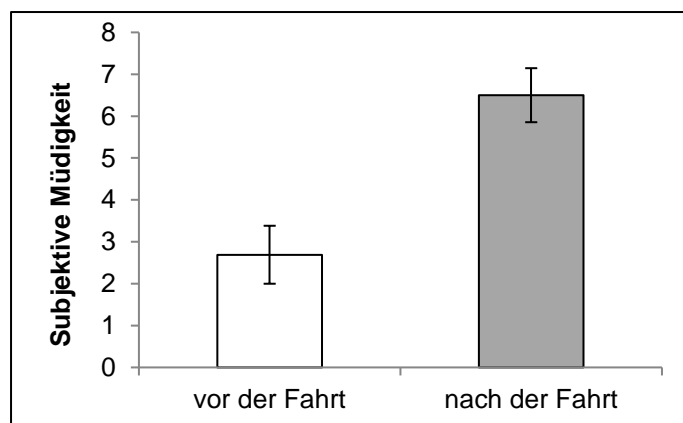


Abbildung 53: Mittlere subjektive Müdigkeit (SOMS) der Probanden abhängig vom Messzeitpunkt. Die Fehlerbalken repräsentieren den Standardfehler.

4.2.1.2 Nutzungsbereitschaft

In Abbildung 54 ist die mittlere Nutzungsbereitschaft der Probanden vor und nach der Versuchsfahrt in Abhängigkeit vom Automatisierungslevel dargestellt. Eine 2×2 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf beiden Faktoren zeigte weder einen Haupteffekt des Faktors „Zeitpunkt“ ($F(1,12) = 3.32, p = .09, \eta_p^2 = .22$) noch einen hypothesenkonformen Haupteffekt des Faktors „Level“ ($F(1, 12) = .60, p = .45, \eta_p^2 = .05$). Allerdings konnte eine signifikante Interaktion beider Faktoren ($F(1, 12) = 4.94, p < .05, \eta_p^2 = .29$) gefunden werden. Demnach bestand nach dem Fahrversuch ausschließlich für Level 3 eine erhöhte Nutzungsbereitschaft.

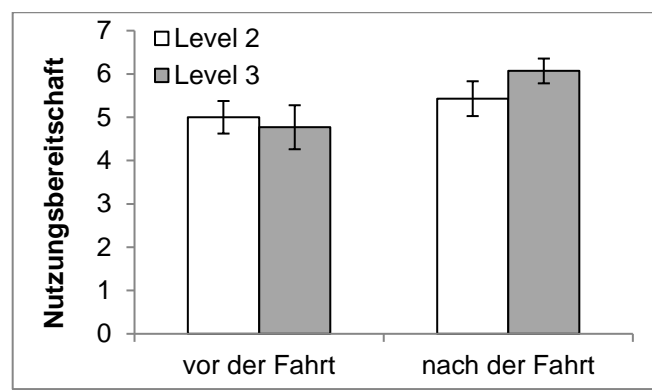


Abbildung 54: Mittlere Nutzungsbereitschaft der Probanden vor und nach der Versuchsfahrt abhängig vom Automatisierungslevel. Die Fehlerbalken repräsentieren den Standardfehler.

4.2.1.3 Vertrauen in die Automatisierungsfunktion

In Abbildung 55 ist das mittlere Vertrauen der Probanden als Summenscore über alle Items der „Automation Trust Scale“ vor und nach der Versuchsfahrt in Abhängigkeit vom Automatisierungslevel dargestellt. Eine 2×2 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf beiden Faktoren zeigte einen Haupteffekt des Faktors „Zeitpunkt“ ($F(1,12) = 31.17, p < .001, \eta_p^2 = .72$). Dementsprechend konnte nach dem Erleben der automatisierten Fahrt ein signifikant höheres Vertrauen als ohne Erfahrung mit der Funktion festgestellt werden. Sowohl der Haupteffekt des Faktors „Level“ ($F(1, 12) = .99, p = .34, \eta_p^2 = .08$) als auch die Interaktion beider Faktoren ($F(1, 12) = 2.31, p = .15, \eta_p^2 = .16$) konnten zufallskritisch nicht abgesichert werden. Tendenziell kann davon ausgegangen werden, dass die Probanden ohne Erfahrung mit der Automatisierung eine höheres Vertrauen in Level 2 als in Level 3 besaßen.

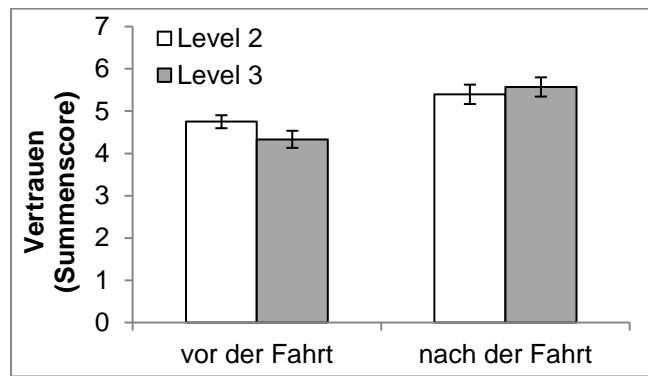


Abbildung 55: Mittlerer Summenscore der „Automation Trust Scale“ als Vertrauensmaß der Probanden vor und nach der Versuchsfahrt abhängig vom Automatisierungslevel. Die Fehlerbalken repräsentieren den Standardfehler.

4.2.1.4 Workload

In Abbildung 56 ist der mittlere Gesamtworkload der Probanden nach der Versuchsfahrt in Abhängigkeit vom Automatisierungslevel und der Übernahme der Fahrzeugsteuerung aus dem jeweiligen Automatisierungslevel dargestellt. Eine 2×2 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf beiden Faktoren zeigte keinen signifikanten Haupteffekt des Faktors „Übernahme“ ($F(1,13) = .01, p = .96, \eta_p^2 < .01$). Sowohl der Haupteffekt des Faktors „Level“ ($F(1, 13) = 8.21, p < .05, \eta_p^2 = .39$) als auch die Interaktion beider Faktoren ($F(1, 13) = 19.06, p < .001, \eta_p^2 = .59$) konnten jedoch zufallskritisch abgesichert werden. Folglich konnte für die Fahrt in Level 2 ein signifikant höherer Workload als in Level 3 nachgewiesen werden. Die Interaktion mit großer Effektstärke lässt sich indes so beschreiben, dass der Workload in Level 2 höher war als bei der anschließenden Übernahme der Fahrzeugsteuerung, hingegen der Workload in Level 3 niedriger war als bei der anschließenden Übernahme der Fahrzeugsteuerung.

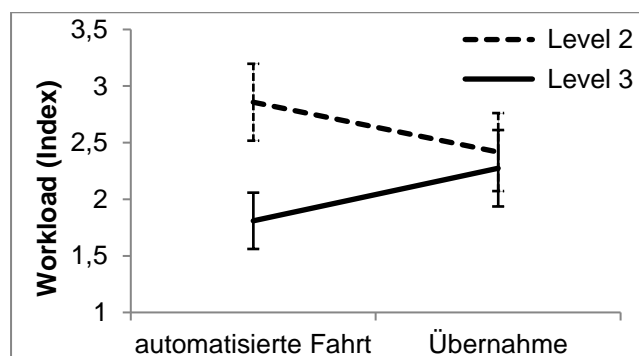


Abbildung 56: Mittlerer Gesamtworkload der Probanden über alle NASA-TLX-Items abhängig von der automatisierten Fahrt oder der Übernahme der Fahrzeugsteuerung im jeweiligen Automatisierungslevel. Ein Index von eins entspricht 10 % des maximal messbaren Workloads. Die Fehlerbalken repräsentieren den Standardfehler.

4.2.1.5 Stresserleben

Abbildung 57 zeigt das mittlere Stresserleben der Probanden als Summenscore über alle Items vor und nach der Versuchsfahrt. Nach der Versuchsfahrt wurde eine Differenzierung zwischen Level 2 und 3 vorgenommen. t -Tests für verbundene Stichproben zeigen an, dass es keine signifikanten Unterschiede im erlebten Stress vor der Versuchsfahrt und nach der Level 2-Fahrt gab ($t(11) = .97, p = .35, \eta_p^2 = .08$). Die Vergleiche zwischen „vor der Versuchsfahrt“ und nach der Level 3-Fahrt ($t(12) = 2.10, p = .06, \eta_p^2 = .27$) sowie zwischen Level 2 und 3 ($t(12) = 1.99, p = .07, \eta_p^2 = .25$) konnten zufallskritisch ebenfalls nicht abgesichert werden, verpassten das Signifikanzniveau jedoch nur knapp. In der Tendenz besteht daher ausschließlich nach Level 3 – entsprechend der Hypothese – eine Stressreduktion verglichen mit dem Versuchsbeginn. Für keine Stressbewertung konnte eine signifikante Korrelation mit dem Persönlichkeitstrait *Neurotizismus* gefunden werden.

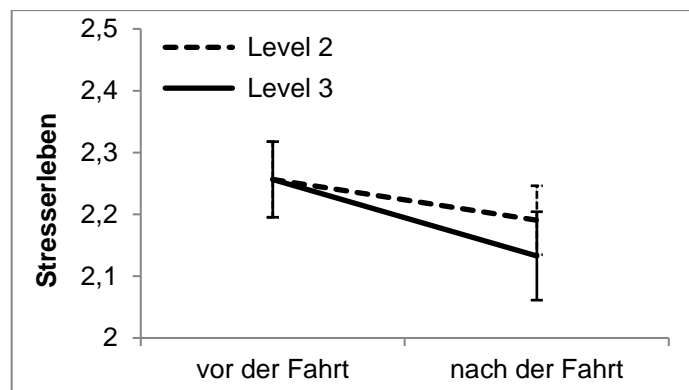


Abbildung 57: Mittlerer Summenscore des SSSQ als Maß des Stresserlebens der Probanden vor und nach der Versuchsfahrt. Nach der Versuchsfahrt wird zwischen Level 2 und 3 unterschieden. Die Fehlerbalken repräsentieren den Standardfehler.

4.2.2 Objektive Metriken

4.2.2.1 EEG-Alpha-Spindelrate: Zeitreihen

Zunächst zeigt Abbildung 58 die an Fahrminute fünf individuell referenzierte Alpha-Spindelrate (relative Alpha-Spindelrate) für jeden Probanden abhängig von der Level 2- oder Level 3-Fahrt über einen gleitenden Mittelwert von 300 s. Insgesamt ist (analog zum ersten Fahrversuch) eine hohe intra- und interindividuelle Varianz der Müdigkeitsverläufe erkennbar. Anders als im ersten Experiment wurde keine Klassifikation von Ermüdung per se vorgenommen, da die Forschungsfragen auf genau diese Unterschiede (hervorgerufen durch verschiedene Automatisierungslevel) abzielten. Besonders auffällig verhalten sich die Müdigkeitsverläufe von

Proband Nr. 10 und Nr. 13. Bei Nr. 10 findet während Level 2 eine starke Ermüdung statt, wohingegen während Level 3 die Ermüdung hypothesenkonform unterdrückt wird. Proband Nr. 13 weist während Level 2 so gut wie keinen Anstieg der Alpha-Spindelrate auf, zeigt jedoch – entgegen der Hypothese – während Level 3 das stärkste Ermüdungsmuster (im Vergleich zu allen anderen Probanden). In den nachfolgenden Analyseschritten soll daher auch untersucht werden, ob womöglich individuelle Einflussfaktoren existieren, die dieses Ermüdungsverhalten beispielhaft erklären könnten.

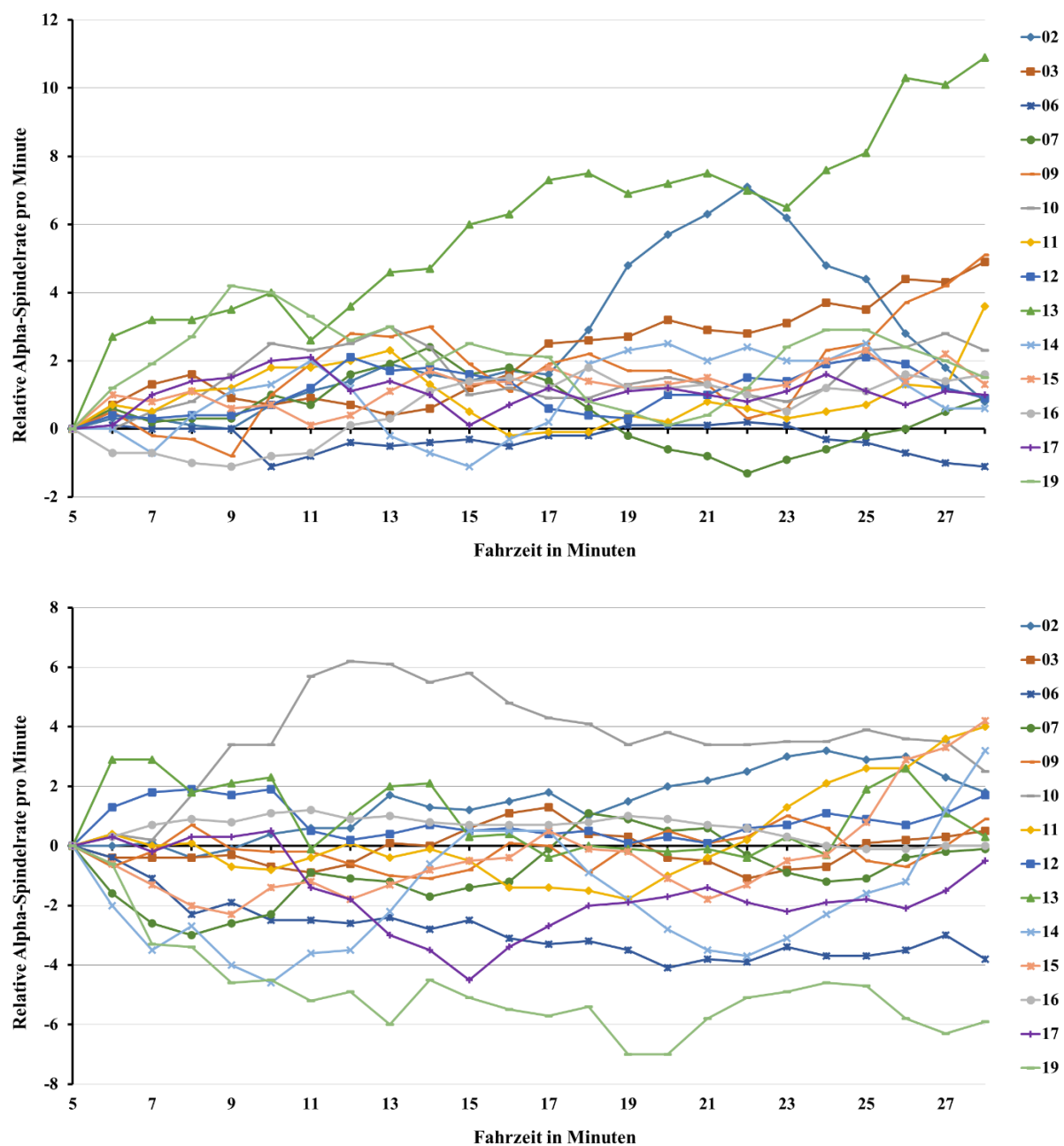


Abbildung 58: Relative EEG-Alpha-Spindelrate pro Minute abhängig von der Fahrzeit (referenziert an Minute 5). Jede Zeitreihe ist einem Probanden zugeordnet. Oben: Level 2. Unten: Level 3.

In Abbildung 59 ist die relative EEG-Alpha-Spindelrate als Differenz zu Minute fünf der automatisierten Fahrt (Level 2 und Level 3) gemittelt über alle Probanden dargestellt. Deskriptiv ist für die Level 2-Fahrt ein Ermüdungsanstieg über die Fahrdauer zu erkennen. Während der Level 3-Fahrt sinkt die Müdigkeit zunächst und verweilt bis zur Minute 25 im Mittel unter dem Referenzwert null. Erst in den darauffolgenden (letzten) drei Minuten ist ein Ermüdungsanstieg festzustellen.

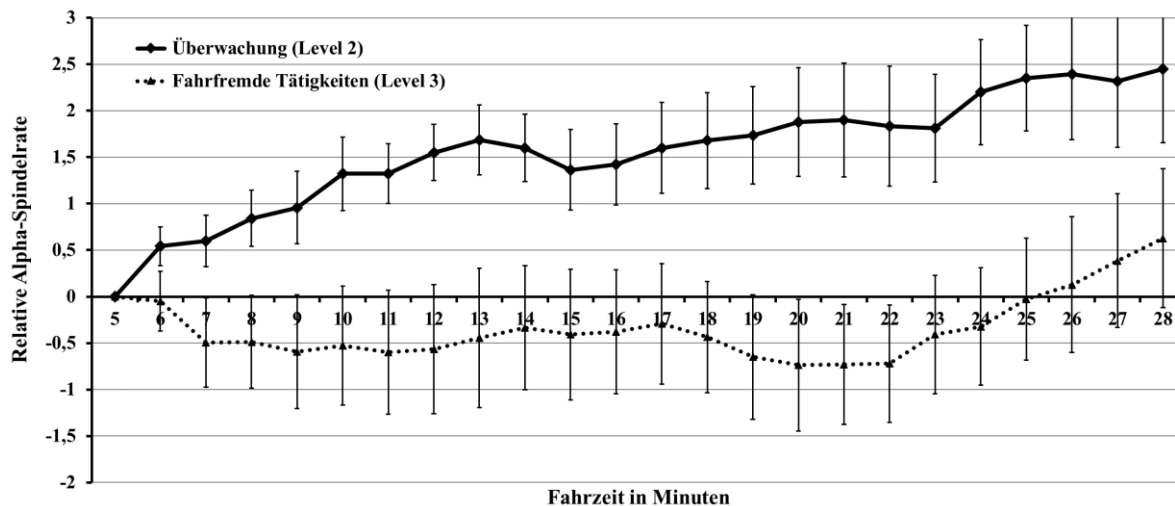


Abbildung 59: Mittlere relative EEG-Alpha-Spindelrate pro Minute abhängig von der Fahrzeit (referenziert an Minute 5) aufgeteilt auf die Level 2- und Level 3-Fahrt. Die Fehlerbalken repräsentieren den Standardfehler.

Für die zufallskritische Analyse wurden zunächst 20 Messwerte von Minute sechs bis einschließlich Minute 25 betrachtet. Eine 2×20 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf beiden Faktoren zeigte entsprechend der Hypothese einen signifikanten Haupteffekt des Faktors „Level“ ($F(1, 13) = 11.14, p < .01, \eta_p^2 = .46$). Demnach konnte über diesen Zeitbereich eine signifikant geringere Ermüdung während Level 3 im Vergleich zu Level 2 ermittelt werden. Post hoc durchgeführte einseitige t -Tests zeigen, dass zu jedem Messzeitpunkt für Level 3 eine geringere relative Alpha-Spindelrate vorlag als für Level 2. Ein signifikanter Haupteffekt des Faktors „Zeitpunkt“ wurde nicht nachgewiesen ($F(19, 247) = 1.42, p = .18, \eta_p^2 = .10$), wohingegen hypothesenkonform ein signifikanter Interaktionseffekt zwischen „Level“ und „Zeitpunkt“ gefunden wurde ($F(19, 247) = 1.67, p < .05, \eta_p^2 = .11$, einseitig getestet). Folglich ermüdeten die Probanden bei Level 2 stärker über die Fahrdauer als dies bei Level 3 der Fall war. Aufgrund der besonders starken Ermüdung von Proband Nr. 13 wurde entschieden, die post hoc t -Tests zusätzlich unter Ausschluss dieses Probanden durchzuführen. Auch hierbei konnte zu jedem Messzeitpunkt in Level 2 eine signifikant höhere Müdigkeit ermittelt werden als in Level 3 (einseitig getestet).

Betrachtet man die letzten drei Minuten automatisierte Fahrt, in der bei Level 3 deskriptiv ein Ermüdungsanstieg (Alpha-Spindelrate > 0) festzustellen ist, so zeigte auch hier eine 2×3 faktorielle Varianzanalyse (ANOVA) mit Messwiederholung auf beiden Faktoren, einen hypothesenkonformen Haupteffekt des Faktors „Level“ ($F(1,13) = 4.99, p < .05, \eta_p^2 = .28$, einseitig getestet). Sowohl der Haupteffekt des Faktors „Zeitpunkt“ ($F(2, 12) = .60, p = .57, \eta_p^2 = .09$) als auch die Interaktion beider Faktoren ($F(2, 12) = .70, p = .51, \eta_p^2 = .11$) konnten zufallskritisch nicht abgesichert werden. Generell wurde für jeden Messzeitpunkt mit Hilfe eines t -Tests für unabhängige Stichproben überprüft, ob womöglich Reihenfolgeeffekte (erst Level 2, dann Level 3 vs. erst Level 3, dann Level 2) die Ergebnisse konfundieren. Für keinen Messzeitpunkt konnte ein signifikanter Reihenfolgeeffekt nachgewiesen werden.

4.2.2.2 Videoanalyse

Level 2

In Abbildung 60 ist die videobasierte Müdigkeitsbeurteilung zu drei verschiedenen einminütigen Messzeitpunkten (t_1 , t_2 und t_3) dargestellt. t -Tests für abhängige Stichproben zeigten, dass das Müdigkeitsurteil zum Zeitpunkt t_2 ($t(12) = -2.37, p < .05, \eta_p^2 = .32$) signifikant und erwartungsgemäß höher ausfiel als zum Messzeitpunkt t_1 . Zum Zeitpunkt t_3 ($t(12) = -2.13, p = .06, \eta_p^2 = .28$) verglichen mit t_1 wurde das Signifikanzniveau knapp verfehlt, sodass hierfür nur eine tendenziell höhere Müdigkeitsbeurteilung berichtet werden kann. Tabelle 17 zeigt die Beurteilungen aufgeteilt auf jeden Probanden; für Proband Nr. 03 lagen keine auswertbaren Daten vor. Es ist erkennbar, dass auch Proband Nr. 13, der den höchsten Anstieg in der relativen Alpha-Spindelrate aufwies, zwischen t_1 und t_2 den größten Sprung in seiner Müdigkeitsbeurteilung besitzt.

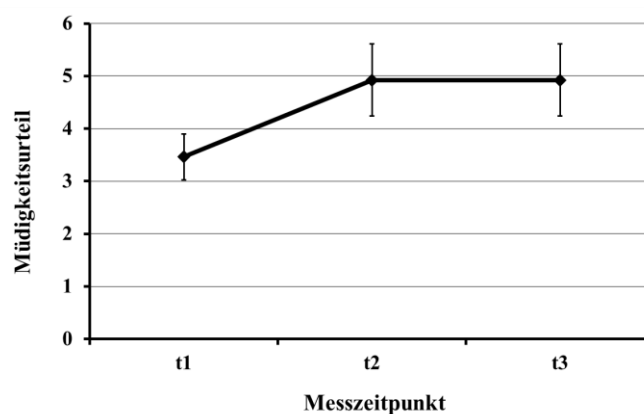


Abbildung 60: Level 2: Mittlere videobasierte Müdigkeitsurteile abhängig von der Fahrdauer zu drei Messzeitpunkten ($t_1 \approx$ Minute 7, $t_2 \approx$ Minute 17, $t_3 \approx$ Minute 27). Die Fehlerbalken repräsentieren den Standardfehler.

Tabelle 17: Level 2: Müdigkeitsurteil für die Messzeitpunkte t_1 , t_2 und t_3 für jeden Probanden.

Proband	t_1	t_2	t_3
02	6	8	8
06	3	2	2
07	5	7	6
09	4	4	4
10	1	2	3
11	4	4	3
12	3	3	2
13	2	8	8
14	2	7	8
15	2	2	3
16	4	8	8
17	3	3	3
19	6	6	6

Level 3

Tabelle 18 zeigt zunächst, welche fahrfremden Tätigkeiten (FFT) welcher Proband wie lange in Summe ausgeführt hat. Ein stetiger Blick durch Front- oder Seitenscheibe des Versuchsfahrzeugs wurde als solcher definiert, wenn er länger als 15 Sekunden andauerte. Blicke, die zwischen einer und 15 Sekunden andauerten, wurden als sog. „Kontrollblicke“ definiert (vgl. Abschnitt 4.1.7). Elf der 14 Probanden beschäftigten sich mindestens einmal mit ihrem Smartphone – dies stellte die häufigste FFT dar; die Tätigkeit „Stricken“ von Proband Nr. 07 ist sicherlich eine Besonderheit. Weiterhin fällt Proband Nr. 10 auf, der die meiste Zeit der Level 3-Fahrt einen stetigen Blick nach draußen zeigte, obwohl gemäß Instruktion sich vorwiegend mit den mitgebrachten oder zur Verfügung gestellten Tätigkeiten (Zeitschriften) beschäftigt werden sollte. Proband Nr. 10 wies zudem die höchste relative Alpha-Spindelrate (vgl. Abbildung 58,

Abschnitt 4.2.2.1) auf. Insgesamt wechselten die meisten Probanden zwischen zwei oder drei Tätigkeiten.

Tabelle 18: Level 3: Art und Dauer der Beschäftigung (in Summe) mit fahrfremden Tätigkeiten (FFT) über die Fahrdauer jedes Probanden.

Proband	stetiger Blick	Smart-phone	Buch / E-Book	Zeitschrift	Essen / Trinken	Stricken	Fahrdauer
02	00:00:21	00:17:46		00:10:25			00:28:32
03		00:00:31		00:28:16			00:28:47
06		00:18:14		00:10:25			00:28:39
07						00:28:51	00:28:51
09		00:26:18		00:02:28			00:28:46
10	00:23:26	00:05:11					00:28:37
11	00:06:06	00:15:50	00:07:30				00:29:26
12		00:28:46					00:28:46
13		00:28:26					00:28:26
14		00:29:03					00:29:03
15			00:20:44	00:08:00			00:28:44
16		00:03:02		00:25:38			00:28:40
17	00:09:40		00:17:26		00:01:09		00:28:15
19		00:28:30					00:28:30

Während der FFT-Beschäftigung wurde das Kontrollblickverhalten der Probanden analysiert. Blicke, die zwischen vier und 15 Sekunden andauerten, wurden als „lang“ zusammengefasst, um genügend Fälle für eine Differenzierung zu kurzen Kontrollblicken (ein bis drei Sekunden) zur Verfügung zu haben. Tabelle 19 zeigt die Anzahl der Kontrollblicke während der Level 3-Fahrt für jeden Probanden. Im Mittel führte die Stichprobe 29 kurze Kontrollblicke ($SE = 6.73$) und elf lange Kontrollblicke ($SE = 2.67$) aus.

Tabelle 19: Level 3: Anzahl von Kontrollblicken entsprechender Dauer während der Beschäftigung mit fahrfremden Tätigkeiten (FFT) für jeden Probanden.

Proband	1 - 3 Sek	4 - 6 Sek	7 - 9 Sek	10 - 15 Sek	Σ_{kurz}	Σ_{lang}
02	19	5	1	1	19	7
03	7				7	0
06	5				5	0
07	16	10	14	4	16	28
09	33	8	7	3	33	18
10	99	19	2		99	21
11	35	7	6	2	35	15
12	40	14	5	3	40	22
13	47	1			47	1
14	19	2	3	2	19	7
15	13	2			13	2
16	21	9	3		21	12
17	51	14	5	4	51	23
19	4				4	0

Weiterhin wurde überprüft, ob die Anzahl kurzer bzw. langer Kontrollblicke mit der mittels Alpha-Spindeln gemessenen Ermüdung während Level 3 oder dem funktionsseitig entgegengebrachten Vertrauen einher geht. Hierfür wurden entsprechende Korrelationen (vgl. Tabelle 20) bestimmt. Demnach waren – entsprechend der Hypothese – signifikant höhere mittlere Alpha-Spindelraten zu beobachten, wenn mehr kurze Kontrollblicke während der FFT-Beschäftigung auftraten. Für lange Kontrollblicke konnte dies nicht nachgewiesen werden. Tendenziell traten auch mehr kurze Kontrollblicke auf, wenn Probanden vor der Fahrt ein geringes Vertrauen in die Level 3-Funktion vermerkten – auch bei einseitiger Testung würde das Signifikanzniveau knapp verpasst werden. Für lange Kontrollblicke konnte auch hier kein Effekt

aufgespürt werden. Das nach der Fahrt erzielte Vertrauen stand in keinem Zusammenhang mit der vorherigen Anzahl kurzer oder langer Kontrollblicke während der FFT-Beschäftigung.

Tabelle 20: Level 3: Korrelationen zwischen der mittleren relativen Alpha-Spindelrate während der Fahrt und dem entgegengebrachten Vertrauen in die Funktion mit der Anzahl von Kontrollblicken während der FFT-Beschäftigung.

abhängige Variable	Kontrollblicke bei FFT	
	kurz	lang
mittlere relative Alpha-Spindelrate (gleitender Mittelwert über 300 s) zwischen Minute 6 und 28	$r = .68$ $p < .01$	$r = .35$ $p = .22$
Vertrauen in Level 3 vor der Fahrt	$r = -.45$ $p = .11$	$r = -.02$ $p = .96$
Vertrauen in Level 3 nach der Fahrt	$r = .24$ $p = .41$	$r = .09$ $p = .76$

Anhand der Videodaten konnte zudem evaluiert werden, ob die Versuchsteilnehmer bei Übernahmeaufforderung erst die FFT beiseitelegten und dann den Lenkradtaster betätigten ($n = 7$), oder umgekehrt erst den Lenkradtaster betätigten (und damit die manuelle Steuerung übernehmen) und anschließend die FFT aus der Hand in das Ablagefach legten ($n = 7$).

4.2.2.3 Übernahme zum manuellen Fahren

In Abbildung 61 sind die Reaktionszeiten (Übernahmezeiten) auf die Übernahmeaufforderung (TOR) abhängig von dem Automatisierungslevel dargestellt, aus dem heraus die Übernahme zum manuellen Fahren erfolgen sollte. Ein t -Test für verbundene Stichproben zeigte einen hypothesenkonformen, signifikanten Effekt zwischen Level 2 und 3 ($t(13) = -5.02$, $p < .001$, $\eta_p^2 = .66$). Nach Level 2 wurde demnach signifikant schneller auf die TOR reagiert als nach Level 3.

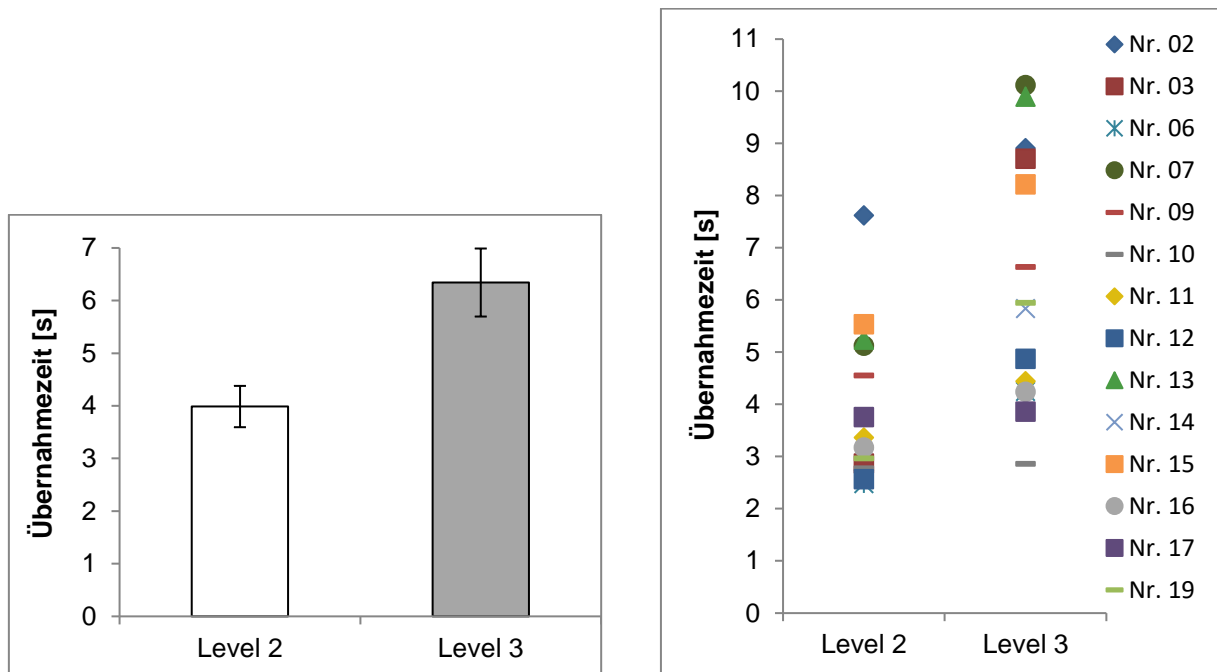


Abbildung 61: Links: Mittlere Übernahmezeit zum manuellen Fahren aller Probanden aus Level 2 und 3. Rechts: Verteilung der Übernahmezeiten für jeden einzelnen Probanden aus Level 2 und Level 3. Die Legende zeigt die Probandennummer. Die Fehlerbalken repräsentieren den Standardfehler.

Es interessierte, welche Faktoren die Übernahmezeit moderieren konnten. Zunächst wurde der Einfluss des Umgangs mit der FFT bei Übernahmeaufforderung nach Level 3 betrachtet, siehe Abbildung 62. Ein t -Test für unabhängige Stichproben zeigte erwartungsgemäß und mit hoher Effektstärke, dass die Übernahmezeit signifikant länger ausfiel, wenn die Versuchsteilnehmer (wie vorgesehen) ihre FFT zunächst beiseitelegten und anschließend den Lenkradtaster betätigten ($t(12) = 5.69, p < .001, \eta_p^2 = .73$).

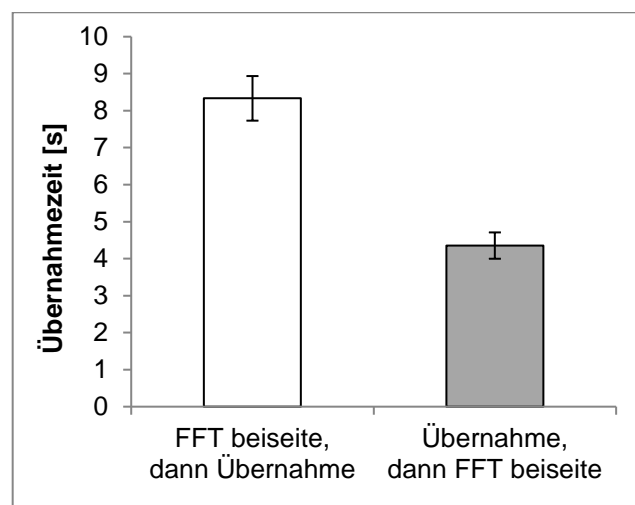


Abbildung 62: Mittlere Übernahmezeit zum manuellen Fahren aus Level 3 abhängig von der FFT-Handhabung bei Übernahmeaufforderung. Die Fehlerbalken repräsentieren den Standardfehler.

Des Weiteren sollte untersucht werden, ob eine Ermüdung während der automatisierten Fahrt eine verlängerte Übernahmezeit zur Folge hatte. Die korrelativen Zusammenhänge sind hierfür in Tabelle 21 aufgeführt. Der physiologische Indikator der mittleren relativen Alpha-Spindelrate im EEG (über die gesamte Fahrdauer bestimmt) konnte eine tendenziell erhöhte Übernahmezeit nach Level 2 mit erhöhter Müdigkeit ausmachen, nicht jedoch nach Level 3. Die in der letzten (28.) Minute automatisierter Fahrt aufgezeichnete Alpha-Spindelrate konnte generell keinen Zusammenhang zwischen Müdigkeit und Übernahmezeit anzeigen. Auffällig – wenn auch ebenfalls bei einseitiger Testung zufallskritisch nicht abzusichern – ist, dass der videobasierte Müdigkeitsindikator (nur während Level 2 vorliegend) auf die größte Tendenz zwischen einer erhöhten Müdigkeit und einer verlängerten Übernahmezeit schließen lässt.

Tabelle 21: Korrelationen zwischen verschiedenen Müdigkeitsindikatoren und der Übernahmezeit nach Level 2 und 3.

Müdigkeitsmessung	Übernahmezeit	
	Level 2	Level 3
relative Alpha-Spindelrate (gleitender Mittelwert über 60 s) zu Minute 28	$r = .16$ $p = .60$	$r = .12$ $p = .68$
mittlere relative Alpha-Spindelrate (gleitender Mittelwert über 300 s) zwischen Minute 6 und 28	$r = .40$ $p = .16$	$r = .00$ $p = 1.00$
videobasiertes Müdigkeitsurteil bei t_3	$r = .48$ $p = .10$	— —

Betrachtet man schließlich die Differenz zwischen der ersten Minute manueller Fahrt nach Übernahme der Fahrzeugsteuerung und der vorherigen letzten (28.) Minute automatisierter Fahrt, so ist entsprechend Abbildung 63 eine Reduktion der relativen Alpha-Spindelrate nach der Übernahme zu erkennen. t -Tests gegen null zeigen hypothesenkonform an, dass sowohl für Level 2 ($t(13) = -1.79$, $p < .05$, $\eta_p^2 = .20$, einseitig getestet) als auch für Level 3 ($t(13) = -3.41$, $p < .01$, $\eta_p^2 = .47$) eine signifikante Aktivierung der Probanden nach Übernahme der Fahrzeugsteuerung zu beobachten ist. Zwischen Level 2 und Level 3 konnte kein signifikanter Aktivierungsunterschied festgestellt werden ($t(13) = 1.46$, $p = .17$, $\eta_p^2 = .14$).

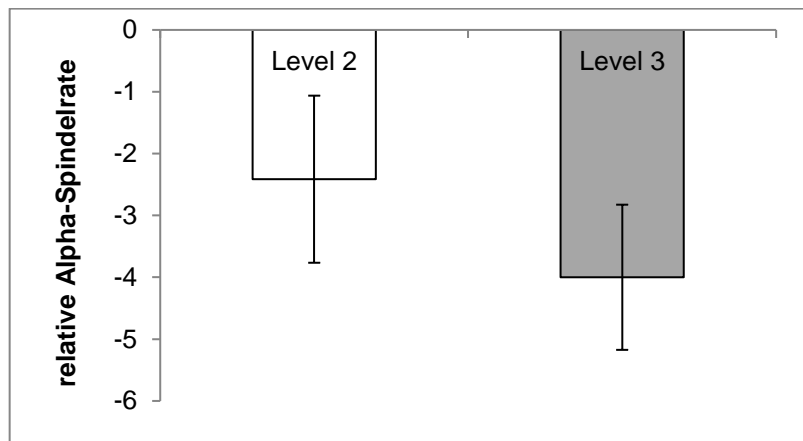


Abbildung 63: Mittlere relative Alpha-Spindelrate als Differenz zwischen erster Minute manueller Fahrt nach Übernahme der Fahrzeugsteuerung und letzter (28.) Minute automatisierter Fahrt. Die Fehlerbalken repräsentieren den Standardfehler.

4.3 Beantwortung der Fragestellungen

Mit Verweis auf Abschnitt 2.5.2 wurden insgesamt fünf Forschungsfragen gestellt, die im Rahmen des zweiten Experiments beantwortet werden sollten.

1. Das Ermüdungsverhalten, das im ersten Experiment während Level 2 nachgewiesen wurde, lässt sich weitestgehend im zweiten Experiment replizieren. Insbesondere die videobasierte Müdigkeitsbeurteilung deutet auf einen hohen Müdigkeitsanstieg in den ersten 17 Minuten hin, anschließend ist der bekannte Sättigungseffekt zu beobachten. Auf Basis der Alpha-Spindel-basierten Analyse zeigte sich ein flacherer Anstieg der Müdigkeit über die gesamte Fahrdauer von 28 Minuten hinweg. Diese Unterschiede zum ersten Experiment sind dadurch zu erklären, dass keine Klassifikation der Versuchsteilnehmer anhand ihrer Spindelraten in eine (vermeintlich) „ermüdete“ und „nicht ermüdete“ Teilstichprobe erfolgte. Der gemessene Workload während der Level 2-Fahrt entsprach ziemlich genau dem Niveau, das auch in Experiment 1 von den Teilnehmern zurückgemeldet wurde. Auch im zweiten Experiment traten immens hohe individuelle Unterschiede im Ermüdungsverhalten auf.
2. Die Nutzung eines Level 3 kann im Vergleich zu einem Level 2 in der Lage sein, Ermüdungserscheinungen über die Fahrdauer vorzubeugen. Bei Bearbeitung fahrfremder Tätigkeiten (FFT) wurde die Ausbildung von Müdigkeit über etwa 25 Minuten erfolgreich unterdrückt, anschließend muss selbst in einem Level 3 – mutmaßlich aufgrund der lang an-

dauernden Bearbeitung der FFT – mit zunehmender Müdigkeit gerechnet werden. Während Level 3 vermerkten die Versuchsteilnehmer eine signifikant geringere Beanspruchung als in der Level 2-Fahrt. Die ermittelten Workload-Niveaus sind gut mit Ergebnissen aus bestehenden Forschungsarbeiten vereinbar. Auch das Stresserleben der Teilnehmer war unter Level 2 signifikant höher als unter Level 3. Damit konnte in Ergänzung zum ersten Experiment gezeigt werden, dass die Überwachung einer automatisierten Fahrfunktion mitunter auch eine Überbeanspruchung bedingen kann, die die Ausbildung passiver, aufgabenbezogener Müdigkeit fördert.

Es wurde erkennbar, dass die Probanden zu Versuchsbeginn einem Level 2 ein höheres Vertrauen entgegenbrachten als einem Level 3. Womöglich besaßen sie bei Abwendung von der Fahraufgabe sicherheitskritische Bedenken. Diese wurden ausgeräumt, sobald sie ihre Erfahrungen mit der Funktion gesammelt hatten. Die Bereitschaft ein Level 3 zu nutzen stieg sogar nach dem Fahrversuch signifikant an.

3. Tatsächlich übernahmen die Probanden nach Level 3 im Mittel langsamer die manuelle Fahrzeugsteuerung, wenn sie ihre FFT beiseitelegten. Sobald aus Level 3 jedoch die Übernahmebereitschaft bestätigt wurde, und die FFT noch in der Hand gehalten wird, waren die Fahrer schneller und benötigen im Schnitt vier Sekunden. Diese vier Sekunden wurden ebenfalls nach Level 2 benötigt – eine in Relation lange Zeitdauer, die sehr wahrscheinlich auf die hohe Müdigkeit zurückzuführen ist: Sowohl für die mittlere relative Alpha-Spindelrate über die gesamte Level 2-Fahrdauer als auch für das videobasierte Müdigkeitsurteil kurz vor Übernahmeaufforderung, konnten hohe positive Korrelationen gefunden werden, deren zufallskritische Absicherung das Signifikanzniveau nur knapp verfehlte. Da die Müdigkeit in Level 3 nur marginal anstieg, konnte in Bezug auf die Übernahme kein Zusammenhang festgestellt werden. Insgesamt wurden die Übernahmen aus Level 2 und Level 3 vergleichbar beanspruchend wahrgenommen.
4. Es wurde erwartet, dass die Übernahme zum manuellen Fahren eine Durchbrechung der Monotonie bewirkt, wodurch die Fahrer aktiviert werden sollten. Dies konnte mit großer Evidenz bestätigt werden. Nach Level 2 reduzierte sich die Müdigkeit im Mittel um die Hälfte, nach Level 3 wurden die Teilnehmer sogar wacher als zu Beginn des Fahrversuchs. In der Tendenz schien der Aufgabenwechsel nach Level 3 demnach eine stärkere Aktivierung der Fahrer hervorzurufen als nach Level 2.

5. In Level 3 müssen nicht zwangsweise Wechsel in der Beschäftigung mit FFT vorliegen, um eine Müdigkeitssuppression zu erreichen. Vielmehr scheint es wichtig zu sein, sich kontinuierlich einer motivierenden Tätigkeit zu widmen. Fahrer, die nur ein geringes Vertrauen in die Level 3-Automatisierung besaßen, führten entsprechend häufiger kurze Kontrollblicke (zwischen einer und drei Sekunden) aus, die mit einer höheren Ermüdung in Zusammenhang standen. Auch dieses Ergebnis steht im Einklang mit der Hypothese, wonach Überwachungstätigkeiten höhere Müdigkeitsniveaus herbeiführen können.

4.4 Diskussion

Der Schwerpunkt des zweiten Experiments lag auf der vergleichenden Analyse der Müdigkeitsentwicklung zwischen einer Level 2- und einer Level 3-Fahrt. Für Level 2 sollte evaluiert werden, ob die Ergebnisse aus dem ersten Experiment repliziert werden können. In Level 3 interessierte besonders die Rolle der fahrfremden Tätigkeiten (FFT): War eine motivierende Beschäftigung in der Lage, Ermüdungserscheinungen vorzubeugen? Diese Aspekte sollen nachfolgend zuerst beleuchtet werden. Anschließend wird die Übernahme der Fahrzeugsteuerung nach beiden Leveln betrachtet – auch hier ist die Abschätzung des Aktivierungspotentials der FFT sowie eine Einordnung der Übernahmezeiten das Ziel. Für eine umfassende Abschätzung des Fahrerzustands werden weitere Einflussfaktoren diskutiert, die auf die Ermüdung während Level 2 und Level 3 einwirken konnten. Schließlich soll auch hier der Untersuchungsansatz des zweiten Experiments rekapituliert werden, um Empfehlungen für zukünftige Forschungsvorhaben aufzuzeigen.

Müdigkeitsverlauf in Level 2 und Level 3 und die Rolle von FFT

Grundsätzlich gaben die Probanden auch im zweiten Experiment nach dem Fahrversuch an, signifikant müder zu sein als vor dem Versuch. Die Müdigkeitszunahme beträgt etwa vier Skalenpunkte auf der SOMS und fällt damit leicht niedriger aus als im ersten Experiment. Allerdings muss hier berücksichtigt werden, dass die Probanden im zweiten Experiment bereits vor Antritt des Fahrversuchs eine leicht höhere Müdigkeit besaßen. Die maximal erreichte subjektive Müdigkeit ist annähernd vergleichbar zum ersten Experiment (marginal niedriger). Letztlich mussten die Teilnehmer nicht ausschließlich in einem Level 2 fahren und überwachen, sondern auch in einem Level 3 FFT nutzen, was eine Aktivierung herbeiführen sollte.

Tatsächlich ist den gemittelten Verläufen der relativen Alpha-Spindelrate über die Fahrdauer zu entnehmen, dass während Level 3 eine Müdigkeitssuppression über ca. 25 Minuten möglich

ist, wohingegen während Level 2 ein monotoner Müdigkeitsanstieg beobachtet werden kann. Verglichen mit dem ersten Experiment verläuft dieser Anstieg flacher und erreicht nicht nach 20 Fahrminuten sein Maximum, sondern nimmt fortwährend bis zur 28. Fahrminute zu. Hierfür lassen sich verschiedene Erklärungen finden. Zum einen fand keine Klassifikation der Versuchsteilnehmer in vermeintlich „ermüdet“ und „nicht ermüdet“ wie im ersten Fahrversuch statt. Grund hierfür war, dass die Suppression von Müdigkeit Teil des Forschungsgegenstands war, und entsprechend der Hypothese durch die Level 3-Fahrt hervorgerufen werden sollte. Insofern ist davon auszugehen, dass ein Anteil der Stichprobe des vorliegenden zweiten Experiments während Level 2 ein Ermüdungsmuster aufwies, welches im ersten Experiment anhand der Hochpassfilter zu einer „nicht ermüdet“-Klassifikation geführt hätte. Betrachtet man Abbildung 58 (vgl. Abschnitt 4.2.2.1), so scheint sich diese Vermutung zu bestätigen. Zum anderen lässt sich nicht mit Bestimmtheit sagen, ob nach der 28. Minute die Spindelrate nicht mehr weiter gestiegen wäre, also auch eine Müdigkeitssättigung stattgefunden hätte. Wenn ja, wäre das Maximum nach der Fahrdauer von 28 Minuten erreicht – leicht verzögert im Vergleich zum ersten Experiment.

Wirft man einen Blick auf die während der Level 2-Fahrt erhobenen Müdigkeitsbeurteilungen auf Basis der Videodaten, so ist genau die aus Experiment 1 bekannte Müdigkeitssättigung wiederzufinden: Zwischen der siebten und der 17. Fahrminute steigt die Müdigkeit signifikant an, und verbleibt äußerst konstant auf einem mittleren Niveau („mittlere Müdigkeit (–)“) bis zum Ende der Versuchsfahrt. Dieses Ergebnis stützt einmal mehr die validen Eigenschaften der Alpha-Spindel-basierten Müdigkeitsdetektion – zumal erkennbar ist, dass beispielsweise ein Proband (Nr. 13) mit hoher Müdigkeitszunahme anhand des Videoratings, ebenfalls im EEG den steilsten Anstieg aufweist. Insgesamt sind die individuellen Unterschiede in den Ermüdungsverläufen, äquivalent zum ersten Experiment, erwartungsgemäß hoch.

Betrachtet man den mittleren Müdigkeitsverlauf während Level 3 genauer, so ist nicht nur eine Müdigkeitssuppression über die Fahrdauer erkennbar, sondern sogar eine Reduktion der Ermüdungsstärke. Bis zur 25. Fahrminute verweilt das Müdigkeitsniveau unterhalb des Ausgangspunkts zur fünften Minute – insofern kann den Probanden sogar eine höhere Wachheit bis zu diesem Zeitpunkt, verglichen mit dem Versuchsbeginn, bescheinigt werden. Nach diesen 25 Minuten steigt auch in Level 3 die Müdigkeit allmählich an, erreicht aber nicht das Niveau in Level 2. Daher kann geschlussfolgert werden, dass das Aktivierungspotential durch die Beschäftigung mit den motivierenden FFT nach dieser Zeit aufgezehrt ist, und selbst motivierende FFT mit der Zeit aufgrund von Monotonie und/oder Langeweile Müdigkeit hervorrufen können.

Viele Versuchsteilnehmer wechselten zwischen zwei oder drei FFT während der Level 3-Fahrt – diese Strategie kann einer Ermüdung, die auf die FFT selbst zurückzuführen ist (vgl. Jarosch et al., 2017, 2019a), vorbeugen. Andersherum könnte man mutmaßen, dass ohne Wechsel eine höhere Müdigkeit vorliegen sollte. Solche Annahmen konnten nur punktuell (aufgrund der Stichprobengröße) geprüft werden: Betrachtet man wieder Proband Nr. 13, der während Level 3 faktisch keine Müdigkeitszunahme zeigt, jedoch in Level 2 am stärksten ermüdet, hat genau dieser Proband nur eine einzige Tätigkeit ausgeführt, nämlich die Nutzung seines Smartphones. Mit Abstand wurde diese Tätigkeit in der Stichprobe am häufigsten gewählt – schließlich ist das Smartphone mittlerweile ständiger Begleiter und besitzt ein breites Nutzungsangebot. Diese Problematik bedingt allerdings, dass nicht mehr differenziert werden konnte, ob sich der Teilnehmer mit dem Smartphone vorwiegend visuell, kognitiv oder motorisch beschäftigte. Beim Lesen einer Zeitschrift sollte z.B. die motorische Modalität (ausgenommen das Umblättern) einen geringeren Anteil haben. Genau solche Ableitungen wären aber wichtig, um entsprechend des Fahrerverfügbarkeitsmodells (Marberger et al., 2018; vgl. Abschnitt 2.4.3) Aussagen über die Passung zwischen dem Fahrerzustand während Level 3 und den nach Übernahme der manuellen Fahrzeugsteuerung benötigten Ressourcen treffen zu können. Diese modalitätsbezogene Fragestellung war nicht Schwerpunkt vorliegenden Experiments und wurde in der jüngeren Vergangenheit bereits intensiv beforscht (siehe z.B. Jarosch et al., 2019).

Interessant für die Fragestellungen vorliegender Arbeit ist umso mehr, dass beispielsweise Proband Nr. 10 einen Großteil der Level 3-Fahrt die Fahrumgebung beobachtete und sich nicht mit einer mitgebrachten Tätigkeit beschäftigte. Dieser Proband wies die höchste Ermüdung auf, obwohl die Level 3-Fahrt diese verhindern sollte. Diese Beobachtung stützt die in Abschnitt 2.4.5.2 thematisierte Problematik, dass eine Level 3-Fahrt nicht per se Müdigkeitseffekten vorbeugen kann – vielmehr ist die Bearbeitung mit FFT entscheidend. Es besteht daher Grund zu der Annahme, dass eine Level 3-Fahrt ohne oder mit Bearbeitung monotoner bzw. langweiliger FFT vergleichbar ermüdend ist, wie eine Level 2-Fahrt. Hierauf deuten auch andere Forschungsarbeiten hin (Feldhütter et al., 2017; Schömig et al., 2015; Vogelpohl et al., 2017; Weinbeer et al., 2017). Da die FFT-Bearbeitung nicht zwangsläufig ein Level 3 definiert, sondern diese lediglich ermöglicht ist, müssen verkehrssicherheitsbezogene Vorteile eines Level 3 grundsätzlich in Bezug zu der FFT-Bearbeitung diskutiert werden. In der Konsequenz wäre nachfolgende Forschung wünschenswert, die mit einem vergleichbaren Versuchsaufbau wie im vorliegenden Experiment (mit objektiver Müdigkeitsmessung in einer Realfahrt), gezielt die fahrerzustandsbezogenen Unterschiede zwischen einem Level 2 und einem Level 3 ohne FFT-Beschäftigung untersucht.

Übernahme zum manuellen Fahren

Nach jeweils 28-minütiger Level 2-Fahrt oder Level 3-Fahrt endete die automatisierte Fahrperiode mit einer Übernahmeaufforderung zum manuellen Fahren. Auf einem geraden Teilstück der Teststrecke sollten die Versuchsteilnehmer ohne zeitliche Kritikalität oder gar drohender Kollisionsgefahr ihre Übernahmebereitschaft mit Druck auf den Lenkradtaster signalisieren. Infolgedessen übergab der Wizard-Fahrer seine Steuerung an den Probanden. Die Übernahmezeit aus Level 2 heraus betrug im Mittel ca. vier Sekunden. Die Definition eines Level 2 sieht allerdings vor, dass der Fahrer unmittelbar, wenn die funktionsseitige Regelung endet, die Fahrzeugsteuerung übernehmen muss. Um dies gewährleisten zu können, muss er permanent die automatisierte Fahrt überwachen. Vor diesem Hintergrund sind die *im Mittel* gemessenen vier Sekunden sehr lang. Zwar übernehmen manche Probanden schneller (Niemand unterhalb von zwei Sekunden), es gibt jedoch auch Teilnehmer, die über fünf Sekunden oder fast acht Sekunden benötigen. Obwohl weder die mittlere relative Alpha-Spindelrate über die gesamte Fahrzeit ($r = .40$, $p = .16$) noch das videobasierte Müdigkeitsurteil kurz vor Übernahmeaufforderung ($r = .48$, $p = .10$) signifikant mit der Übernahmezeit korrelierte, weisen doch teils hohe Korrelationen darauf hin, dass mit höherer Müdigkeit eine längere Übernahmezeit nach Level 2 einherging. Aus Perspektive der Verkehrssicherheit ist diese Feststellung kritisch – schließlich hat im vorliegenden Fall das Versuchsfahrzeug bei ca. 70 km/h in vier Sekunden knapp 80 m zurückgelegt. Ebenso unerwartet wie im vorliegenden Experiment könnte im realen Straßenverkehr eine Level 2-Funktion eine Systemgrenze erreichen und damit ihre Regelungstätigkeit beenden. Häufig sind dies Verkehrssituationen, die über eine gewisse Komplexität verfügen – schließlich konnte die Funktion sie selbst nicht mehr lösen. Legte dann der Fahrer ca. 80 m aufgrund seiner Ermüdung faktisch ohne Regelungstätigkeit zurück (auf einer Bundesautobahn ggf. eine noch längere Wegstrecke, weil die Fahrgeschwindigkeit höher ist), kann die Unfallgefahr steigen.

Nach der Level 3-Fahrt waren die Probanden im Experiment im Mittel genauso schnell bereit, die Fahrzeugsteuerung zu übernehmen, wie nach Level 2, jedoch nur, wenn sie ihre FFT in der Hand behielten. Die Instruktion sah jedoch explizit vor, diese abzulegen – schließlich sollte beispielsweise keine manuelle Weiterfahrt mit gleichzeitigem Halten eines Smartphones erfolgen. Hier stellt sich aber die Frage, warum die Hälfte der Stichprobe trotzdem erst die Fahrzeugsteuerung übernehmen wollte, und erst anschließend die FFT beiseitelegte. Eine Erklärung könnte darin liegen, dass die Teilnehmer keine intensive Erfahrung in der Übernahme der Fahrzeugsteuerung besaßen und schnellstmöglich (i.S. eines Reflexes auf Ton und Icon im Kombiinstrument) den Lenkradtaster betätigten.

Wenn die Versuchsteilnehmer zunächst ihre Tätigkeit in das Ablagefach auf dem Beifahrerplatz legten, lag ihre Übernahmezeit im Mittel über acht Sekunden. Verglichen mit anderen Studienergebnissen (für einen Überblick siehe Zhang et al., 2019), umfassen sowohl die vier Sekunden ohne das Beiseitelegen der FFT als auch die acht Sekunden mit Beiseitelegen eine verhältnismäßig lange Zeitdauer. Auch hier lässt sich der Ermüdungsfaktor als eine Erklärung anführen: Nach etwa 25 Fahrminuten steigt auch in Level 3 die Müdigkeit allmählich an. Dieser Fahrerzustand kann die Übernahmezeiten verlängert haben – dies sei aber nur mit Vorsicht interpretiert, da keine unmittelbaren Zusammenhänge zwischen einem Alpha-Spindel-basierten Maß und der Übernahmezeit gefunden werden konnten. Eine weitere Erklärung kann in der Unterbrechbarkeit der FFT gefunden werden: Falls zum Zeitpunkt der Übernahmeaufforderung eine schlechte Unterbrechbarkeit der FFT vorlag – dies wurde zwar nicht kontrolliert, könnte aufgrund des hohen motivationalen Charakters der FFT jedoch anzunehmen sein – wird die FFT ggf. nur ungern oder nur verzögert aus der Hand gelegt.

Schließlich wurde die Frage gestellt, ob die Übernahmeaufforderung eine Aktivierung im Fahrerzustand hervorruft. Betrachtet man die Differenz der Alpha-Spindelrate zwischen der letzten, also 28. Minute automatisierter Fahrt und der ersten Minute manueller Fahrt, so zeigten sich deutliche Aktivierungsmuster. Nach Übernahme aus Level 2 war eine im Mittel um mehr als zwei Spindeln pro Minute reduzierte Rate messbar, aus Level 3 betrug die Reduktion sogar ca. vier Spindeln pro Minute. Damit verringerte sich die Müdigkeit nach Level 2 um ungefähr die Hälfte, nach Level 3 wurden die Probanden noch wacher als zu Versuchsbeginn. Obwohl die Varianz zwischen den Probanden zu groß war, um die Unterschiede zwischen Level 2 und 3 zufallskritisch abzusichern, soll dennoch die sich (entgegen der Hypothese) abzeichnende Tendenz diskutiert werden: Die Aktivierung nach Level 3 war höher als nach Level 2, obwohl die Müdigkeit während Level 3 niedriger war als während Level 2. Kann der jeweilige Aufgabenwechsel für dieses Phänomen die entscheidende Rolle gespielt haben?

Mit Beendigung der Level 2-Fahrt wurde manuell gefahren und somit die Überwachungstätigkeit eingestellt, hingegen musste nach Level 3 die FFT-Beschäftigung beendet werden. In letzterem Punkt könnte eine Erklärung zu finden sein, warum nach Level 3 mehr Aktivierung zu beobachten ist, als nach Level 2. Womöglich ist nämlich der Aufgabenwechsel, der mit Ausführung der manuellen Steuerung nach der automatisierten Fahrt stattfindet, hinsichtlich der Ressourcenbeanspruchung zwischen den Leveln zu unterscheiden: Die Ressourcen, die nach Level 2 zwischen „Überwachung“ und „manuellem Fahren“ beansprucht werden, sind ggf. ähnlicher als nach Level 3. Aus Level 3 muss eine fahrfremde Tätigkeit beendet werden, diese unter motorischer Beanspruchung auf den Beifahrerplatz abgelegt, und anschließend die manuelle

Fahraufgabe ausgeführt werden. So könnte gemutmaßt werden, dass das Aktivierungspotential dann am größten ist, wenn die Aufgaben, zwischen denen gewechselt werden soll, hinsichtlich ihrer Ressourcenbeanspruchung möglichst verschieden sind. Diese Überlegung ist auf den ersten Blick nur schwer mit dem Fahrerverfügbarkeitsmodell (Marberger et al., 2018; vgl. Abschnitt 2.4.3) vereinbar, da der Fahrerausgangszustand und der Fahrerzielzustand weit auseinander lägen. Allerdings postuliert das Modell nicht nur, dass eine der Ziel-Fahraufgabe nahe Ausgangsaufgabe wünschenswert wäre, um einen adäquaten Zielzustand zu erreichen, sondern auch, dass der Ausgangszustand im Hinblick auf den Zielzustand eine Anpassung erfahren sollte, wenn Abweichungen untereinander vorliegen. Folglich könnte zusätzlich eine Anpassung des im Modell verankerten „Arousal-Levels“ durch *Aktivierung* während der Übernahme-phase (im Modell der „Driver state transition process“) stattfinden. Berücksichtigt man also vorliegende Forschungsergebnisse, so scheint ein Aufgabenwechsel von Level 3 zum manuellen Fahren die entsprechende Aktivierung gewährleisten zu können, die eine notwendige, aber nicht die hinreichende Bedingung für einen adäquaten Zielzustand ausmacht. Schließlich ist die Frage nach dem notwendigen Situationsbewusstsein dann noch immer nicht geklärt. Hier empfiehlt es sich folglich, sich näher mit den subjektiven Angaben, insbesondere zum bewerteten Workload auseinanderzusetzen.

Müdigkeit in Level 2 und Level 3 – weitere Einflussfaktoren

Auf Basis des ermittelten Workloads im zweiten Experiment lassen sich verschiedene Ableitungen tätigen. Hier ist wichtig, zwischen dem Workload zu unterscheiden, den die Versuchsteilnehmer rückblickend während der Level 2- bzw. Level 3-Fahrt bewerteten und dem Workload, der während der Übernahme zum manuellen Fahren beurteilt wurde. In der automatisierten Fahrperiode konnte für Level 2 eine signifikant höhere Beanspruchung festgestellt werden als für Level 3. Während bei Level 2 der Workload bei ca. 29 % des mittels NASA-TLX maximal messbaren Workloads lag, gaben die Versuchsteilnehmer für Level 3 lediglich ca. 18 % an. Demnach war für Level 2 eine Replikation des ermittelten Workloads aus dem ersten Experiment sehr gut möglich. Die Werte aus Level 3 decken sich ebenfalls gut mit der von De Winter et al. (2014) evaluierten Marke von 22.7 %. Unter Berücksichtigung, dass die Autoren Level 3 sowohl mit als auch ohne FFT in die Zusammenstellung aufnahmen, und in einem Level 3 mit FFT ein niedrigerer Workload als ohne FFT vorliegen sollte, lässt sich der Wert von 18 % sehr gut integrieren. Folglich konnte die Hypothese bestätigt werden, wonach in Level 3 ein geringerer Workload aufgrund geringerer Ermüdung vorherrschen sollte.

Interessanterweise war im ersten Experiment aufgrund eines niedrigeren gemessenen Workloads in Level 2 als üblicherweise beim manuellen Fahren davon ausgegangen worden, dass vorwiegend eine Unterbeanspruchung durch die permanente Überwachungstätigkeit stattgefunden hat. Betrachtet man jetzt die Relation zwischen Level 2 und 3, möchte man mutmaßen, dass in Level 2 ein zu hoher Workload verglichen mit Level 3 vorlag, weswegen auf eine Überbeanspruchung während Level 2 geschlossen werden könnte. Hieran zeigt sich, dass der Dualismus zwischen Über- und Unterbeanspruchung bei Level 2-Nutzung (i.S. einer Vigilanzaufgabe) sehr präsent ist. Insbesondere das zweite Experiment erlaubte es den Probanden, zwischen Level 2 und Level 3 unmittelbar zu vergleichen, weil sie beide Funktionen nacheinander erlebt hatten. Daher kann dem Überbeanspruchungscharakter eines Level 2 im vorliegenden Fall die notwendige Evidenz zugesprochen werden.

Die Übernahme der manuellen Fahrzeugsteuerung war nach Level 2 oder Level 3 vergleichbar beanspruchend. Dies ist insofern bemerkenswert, da – wie zuvor geschildert – nach Level 3 eine komplexere Aufgabe dem Probanden auferlegt wird (insbesondere die Beendigung der FFT und das Beiseitelegen) als nach Level 2. Eine Erklärung könnte sein, dass in die Bewertung des Workloads auch das Aktivierungspotential genau dieses komplexeren Aufgabenwechsels mit einbezogen wurde, sodass im Mittel eine äquivalente Beanspruchung zwischen Level 2 und 3 vorlag. Dies würde die zuvor diskutierte höhere Aktivierung nach Level 3 unterstützen.

Zusätzlich zum Workload sollte das bewertete Stresserleben während Level 2 und Level 3 betrachtet werden. Hier zeigte sich zunächst, dass die Teilnehmer vor der Versuchsfahrt tendenziell ein höheres Stresserleben als nach der Fahrt besaßen – dies mag der Ungewissheit geschuldet sein, was sie wohl in ihrer erstmaligen automatisierten Fahrt erwarten würde. Nach der Fahrt zeichnete sich deutlich ab, dass das Fahren in einem Level 2 mit einem höheren Stress verbunden war als in einem Level 3 – das Signifikanzniveau wurde bei zweiseitiger Testung nur knapp verfehlt. Dieses Ergebnis steht demnach im Einklang mit dem zuvor diskutierten Workload und der von Vigilanzaufgaben bekannten Typologie: Entsprechend des angepassten Müdigkeitsmodells für Level 2 (vgl. Abschnitt 2.3.4) bedingten ein hoher Workload und ein hohes Stresserleben die Ausbildung passiver aufgabenbezogener Müdigkeit. Level 3 kann also unter Berücksichtigung des Faktors „Stress“ einer Ermüdung vorbeugen. Aus diesen Gründen mag auch die Nutzungsbereitschaft nach dem Erleben der Funktionen nur für Level 3 höher sein als zu Beginn des Fahrversuchs. Ob lediglich die Befreiung von der Überwachungsaufgabe und die Freiheit, die eigens mitgebrachten FFT verwenden zu können, eine höhere Nutzungsbereitschaft bedingte, oder auch das damit verbundene geringere Ermüden positiv bewertet wurde, kann nicht zweifelsfrei interpretiert werden.

Schließlich soll das Vertrauen in die Automatisierungsfunktionen diskutiert werden, das einerseits über subjektive Abfragen gemessen und andererseits anhand der ausgeführten Kontrollblicke objektiviert wurde. Auf Basis der Funktionsbeschreibungen, die den Probanden vor dem Fahrversuch für ein Vertrauensurteil für Level 2 und 3 zur Verfügung standen, ist tendenziell ersichtlich, dass die Bearbeitung von FFT während der Fahrt unter Umständen Bedenken hervorruft, wohingegen eine permanente Überwachung (also alles „im Blick zu haben“) mit einem höheren Funktionsvertrauen einhergeht. Da die Stichprobe erstmals eine automatisierte Fahrt auf dem entsprechenden Level erlebte, waren die Teilnehmer womöglich verunsichert, ob die Level 3-Funktion wirklich so fehlerfrei arbeitet, dass man ihr bedenkenlos vertrauen könne. Ihre kurzen Kontrollblicke untermauern diese Feststellung: Es ist eine deutliche Tendenz erkennbar, wonach Probanden mit einem geringeren Vertrauen in die Level 3-Funktion vor der Fahrt mehr kurze Kontrollblicke (zwischen einer und drei Sekunden) während Level 3 durchführten ($r = -.45, p = .11$) als Probanden mit einem höheren Vertrauen. Diese Sicherheitsbedenken zerstreuen sich, nachdem beide Funktionen erlebt wurden – das Vertrauen nach dem Fahrversuch stieg signifikant und auch das Vertrauen in Level 3 glich sich dem in Level 2 an. Es gab jedoch keinen fundierten Anhaltspunkt, wonach eine höhere Anzahl kurzer Kontrollblicke während Level 3 mit einem höheren Vertrauen nach dem Versuch in Zusammenhang gestanden hätte ($r = .24, p = .41$). Generell stellten lange Kontrollblicke (zwischen vier und 15 Sekunden) keinen validen Vertrauensindikator dar. Die Ergebnisse zeigen jedoch, dass (gegeben eine verkehrssichere Level 3-Funktion liegt vor) mittels Exposition und Fahrerfahrung auch gesellschaftliche Vorbehalte in Bezug auf die Verkehrssicherheit höherer Automatisierungslevel ausgeräumt werden können.

Insbesondere interessierte der Zusammenhang zwischen dem Kontrollblickverhalten und der Ermüdung der Stichprobe während Level 3. Die Anzahl kurzer Kontrollblicke korrelierte signifikant mit der mittleren relativen Alpha-Spindelrate über die gesamte Fahrdauer ($r = .68, p < .01$). Probanden, die während Level 3 eine höhere Müdigkeit besaßen, führten also mehr kurze Kontrollblicke aus. Wie zuvor diskutiert, bedingten mehr kurze Kontrollblicke ein niedriges Vertrauen in die Fahrfunktion. Selbstverständlich sind aus diesen Zusammenhängen keine Kausalitäten abzuleiten. Würde man dennoch transitiv argumentieren, ergäbe sich, dass eine höhere Müdigkeit durch niedriges Funktionsvertrauen angezeigt ist.

Bereits im ersten Experiment wurde erwartet, dass eine höhere Müdigkeit mit höherem Vertrauen einhergehen sollte, wenn Fahrer eine hohe Langeweile besaßen. Diese Vermutung galt – aufgrund der anzunehmenden Unterbeanspruchung – für die Level 2-Nutzung, was jedoch

nicht nachgewiesen werden konnte. Dass sich in Level 3 genau das umgekehrte Muster abzeichnete, also Probanden, die ein niedriges Funktionsvertrauen besaßen, stärker ermüden, kann als Hinweis auf das Müdigkeitssuppressionspotential eines Level 3 gewertet werden: Wenn die Versuchsteilnehmer ohne Kontrollblicke auszuführen (weil sie der Funktion vertrauen), ihren motivierenden FFT nachgehen, kann dadurch erfolgreich einer Ermüdung vorgebeugt werden. Sobald die Fahrer jedoch beginnen, „Überwachungsverhalten“ in Form kurzer Kontrollblicke zu zeigen, steigt die Gefahr zu ermüden wieder an. Lange Kontrollblicke schienen zur Aufdeckung dieses Effekts mit insgesamt geringerer Sensitivität in der Lage zu sein, womöglich weil diese seltener ausgeführt wurden. Auch dieses Ergebnis besitzt praktische Relevanz für die Verkehrssicherheit: Wenn Nutzer eines Level 3 ein hohes Funktionsvertrauen besitzen, und ihren (vermeintlich) motivierenden FFT nachgehen, sinkt die Wahrscheinlichkeit zu ermüden. In der Konsequenz besäßen die Fahrer dann bei Übernahme der manuellen Fahrzeugsteuerung einen vorteilhafteren Zustand.

Untersuchungsansatz und Versuchskonzeption

In Summe lässt sich festhalten, dass die Forschungsfragen des zweiten Experiments zufriedenstellend beantwortet werden konnten – auch über die primären Forschungsfragen hinaus wurden wertvolle Erkenntnisse zusammengetragen. Es konnte insbesondere von einer hohen Teststärke aufgrund des Messwiederholungsdesigns profitiert werden. Letztlich waren den Probanden auf diese Weise auch direkte Vergleiche zwischen Level 2 und Level 3 möglich. Prinzipiell erlaubte das Design auch, eine reduzierte Anzahl an Teilnehmern zu untersuchen – dies war insofern wünschenswert, als dass auch die zweite Studie unter hohem Aufwand durchgeführt wurde und mit hohen monetären Kosten verbunden war. Trotzdem wären mit Hilfe eines größeren Stichprobenumfangs manche Analysen höchstwahrscheinlich zufallskritisch abzusichern gewesen.

Es zeigte sich erneut, dass der verwendete „Wizard-of-Oz-Ansatz“ sehr gewinnbringend eingesetzt werden kann, da im vorliegenden Experiment kein einziger Teilnehmer Mutmaßungen über den Wizard-Fahrer äußerte. Allerdings müssen Restriktionen in Bezug auf eine realitäts- bzw. „seriennahe“ Transition zwischen automatisierter und manueller Steuerung berücksichtigt werden: Sobald der Fahrer die Steuerung übernehmen möchte, muss der Wizard-Fahrer diese erst wieder frei geben. Ein spontanes „Abschalten“ z.B. durch Betätigung des Bremspedals (bei gleichzeitiger Handhaltung am Lenkrad), kann aus Sicherheitsgründen nicht realisiert werden. In vorliegender Studie spielte dies keine Rolle – hier war die Bestätigung der Übernahmebe-

reitschaft des Fahrers von Interesse, um Rückschlüsse auf seinen Fahrerzustand ziehen zu können. Wenn tatsächlich spontane Eingriffe in die automatisierte Fahrt untersucht werden, wird der Einsatz eines Fahrsimulators womöglich die bessere Wahl sein.

Bei Betrachtung der „reflexartigen“ Übernahmen nach Level 3, in denen die Fahrer ihre FFT noch in der Hand hielten, wird einmal mehr deutlich, dass dem Design der Mensch-Maschine-Schnittstelle eine entscheidende Rolle in zukünftigen Level 3-Fahrzeugen zukommen wird. Hier muss eine Möglichkeit gefunden werden, einen salienten und der Dringlichkeit angepassten Hinweis auf eine bevorstehende Übernahme zu präsentieren, ohne eine Schreckreaktion hervorzurufen. Nicht zuletzt wäre weiterführende Forschung wichtig, die mit einem ähnlichem Untersuchungsansatz die Rolle der Monotonie, im Speziellen die Verkehrsumgebung und die Verkehrsdichte, auf den Fahrerzustand in Level 2 und Level 3 untersucht. Schließlich werden zukünftig leistungsfähigere Funktionen in die Fahrzeugflotte Einzug erhalten, die auch in komplexeren Verkehrssituationen langandauernd aktiviert bleiben können.

Abschließend empfiehlt es sich, das gewählte Untersuchungsszenario kritisch zu hinterfragen. Nicht nur in Experiment 2, sondern bereits in Experiment 1, fuhren alle Probanden unter höchst kontrollierten Bedingungen in einem (für die Teilnehmer fremden) Versuchsfahrzeug ohne die Involviertheit anderer Verkehrsteilnehmer auf der Ovalbahn einer abgesperrten Teststrecke. Die Vorteile dieser validen Versuchskonzeption im Hinblick auf den Untersuchungsgegenstand wurden zuvor von verschiedenen Perspektiven betrachtet. So wurde im Anschluss an das erste Experiment diskutiert, dass auf diese Weise eine monotone Fahrsituation, wie sie bei Level 2-Nutzung mutmaßlich nicht unüblich ist, nachgebildet werden konnte. Dennoch bleibt die artifizielle Versuchskonzeption als offenkundiger Kritikpunkt bestehen: Insbesondere kann im zweiten Experiment nicht ausgeschlossen werden, dass die Versuchsteilnehmer aufgrund der Beobachtung durch Videokameras und durch das Tragen eines EEG ihr Verhalten in Bezug auf die FFT-Nutzung angepasst haben. Speziell kann das Kontrollblickverhalten der Probanden durch das Untersuchungsszenario beeinflusst worden sein. So könnten im realen Straßenverkehr mitunter mehr Kontrollblicke, ggf. auch mehr längere Kontrollblicke auftreten, weil die Verkehrssituation dies aufgrund ihrer Komplexität veranlasst. In jedem Fall werden aufgrund der hohen Monotonie der Teststrecke die Ermüdungserscheinungen schneller aufgetreten sein als im realen Straßenverkehr bei moderater Verkehrsdichte. Allerdings interessierte speziell im zweiten Experiment der Vergleich zwischen Level 2 und 3, sodass die Relation im Ermüdungsverhalten auch unter artifiziellen Bedingungen konstant bleiben sollte. Zusätzlich könnte das künstliche Szenario die subjektiven Einschätzungen der Probanden zu ihrer Beanspruchung

konfundiert haben; das langandauernde Tragen der EEG-Haube hat womöglich zusätzlich belastet und damit die Beanspruchung gesteigert. Aber auch hier gilt, dass der direkte Vergleich zwischen den Leveln davon nicht berührt sein sollte. Daher wird insgesamt angenommen, dass die Entscheidung für eine hohe Kontrollierbarkeit im Versuch mit mehr Validitätsvorteilen für die Beantwortung der Fragestellungen verbunden war als Validitätsabstriche dabei in Kauf genommen werden mussten. Wenn dennoch objektive Müdigkeitsmessungen unter Verwendung von Automatisierungsfunktionen im realen Straßenverkehr zukünftig durchgeführt werden sollten, muss ein entsprechendes Sicherheitskonzept ausgearbeitet sein – schließlich läge dann die Fahrverantwortung bei einem Probanden mit einem für die Fahraufgabe mutmaßlich unzureichenden Zustand. Die Resultate einer solchen Untersuchung könnten jedoch helfen, das tatsächliche Übertragbarkeitspotenzial vorliegender Ergebnisse auf reale Gegebenheiten fundierter abzuschätzen.

Die zentralen Ergebnisse beider Experimente aus vorliegender Arbeit sollen nun in einem gemeinsamen Fazit zusammengefasst werden.

5 Fazit

Die vorliegende Arbeit konnte neue Erkenntnisse in Bezug auf den Fahrerzustand beim automatisierten Fahren zusammentragen. Unter hoher externer wie interner Validität war es möglich, Einblicke in die fahrerseitige Müdigkeitsentwicklung zu erhalten, die in diesem Detaillierungsgrad bis dato noch nicht ermittelt wurden. Insbesondere der Einsatz eines Wizard-of-Oz-Fahrzeugs auf einer Teststrecke erlaubte es, eine realitätsnahe Nachbildung von Automatisierungsfunktionen gemäß eines SAE Level 2 und Level 3 zu erzeugen und diese unter kontrollierten, experimentellen Bedingungen den Versuchsteilnehmern erlebbar zu machen. Auf diese Weise konnten in einer monotonen Umgebung Auswirkungen einer langandauernden Nutzung dieser Funktionen auf den Fahrerzustand – konkret die Fahrermüdigkeit – objektiv untersucht werden. Die Müdigkeitsmessungen wurden durch immense intra- und interindividuelle Varianzen erschwert, die jedoch teilweise anhand verschiedener Einflussfaktoren erklärt werden konnten. Beispielsweise bedingte das Alter der Versuchsteilnehmer und ihr Koffeinkonsum vor dem Fahrversuch erheblich, wie groß die Müdigkeitszunahme über die Fahrdauer war. So ermüdeten ältere Probanden und Teilnehmer unter Koffeineinfluss signifikant weniger.

Die EEG-basierte Müdigkeitsdetektion über Alpha-Spindeln erwies sich als weitestgehend valider Indikator. Allerdings zeigte sich, dass die Spindelrate, weil diese individuell referenziert wurde, vielmehr das Ermüdungspotential über die Fahrdauer abbilden kann als absolute Müdigkeitswerte zu liefern. Letzteres wurde aber durch das Videorating ermöglicht. Beide Indikatoren deuteten darauf hin, dass innerhalb der ersten 20 Minuten einer Level 2-Fahrt die Ermüdung monoton ansteigt und anschließend auf einem mittleren Niveau verbleibt. Hier kann von einer Müdigkeitssättigung gesprochen werden. Auch die subjektiven Urteile vor und nach der 60-minütigen Level 2-Fahrt gaben Hinweise auf einen deutlichen Müdigkeitsanstieg über die Versuchsdauer. Insbesondere waren genau diese Selbstbewertungen vor dem Versuch in der Lage, das Ermüdungspotential, welches dann über die Alpha-Spindel-Analyse sichtbar wurde, zu präzisieren. Je tendenziell niedriger die Müdigkeit vor dem Versuch, desto größer war das Steigerungspotential. Insofern muss festgehalten werden, dass alle drei in vorliegender Arbeit verwendeten Müdigkeitsindikatoren sich in ihrer Spezifität ergänzten und ihren Beitrag für eine möglichst umfassende Fahrerzustandsbestimmung geleistet haben.

Einer geringeren Funktionszuverlässigkeit (Automatisierungsfehler alle ca. acht Minuten) wurde von den Versuchsteilnehmern eine erhöhte Nutzungsbereitschaft attestiert. Tatsächlich waren diese Fehler in der Längsführung und als Querablage während Level 2 in der Lage,

Müdigkeit kurzzeitig (für ca. zwei Minuten) zu reduzieren. Hierbei sind aber verschiedene Rahmenbedingungen zu beachten: Einerseits musste die Müdigkeit ihr maximales Niveau im Fahrversuch erreicht haben, damit diese Fehler überhaupt wirksam sein konnten. Andererseits hatte eine häufige und regelmäßige Präsentation der Fehler zur Folge, dass sich ein Habituationseffekt über die Fahrdauer einstellte, sodass sich nach sechs Fehlern und etwa 50 Minuten Fahrt, keine spürbare Aktivierung mehr abzeichnete. Dieses Ergebnis wurde auch anhand der Reaktionszeit auf die Querablage nach 50 Fahrminuten deutlich. Probanden, die schon sechs Fehler erlebt hatten, reagierten im Mittel ca. eine Sekunde *bevor* das Versuchsfahrzeug seinen Fahrstreifen verlassen hatte – es fand also eine Antizipation statt. Hingegen reagierten „ermüdete“ Teilnehmer, die erstmalig eine Fehlerpräsentation erlebten, erst ca. eine halbe Sekunde *nachdem* der Fahrstreifen bereits verlassen war. Damit wird deutlich, mit welchen fahrerzustandsbezogenen Einbußen während einer langen Level 2-Fahrt, aufgrund von Ermüdung, zu rechnen ist. Aus Verkehrssicherheitsperspektive können in diesem Zusammenhang kritische Situationen entstehen – in welcher Weise die Adäquatheit der Reaktionsauswahl in einer solchen Verkehrssituation ebenfalls durch Müdigkeit beeinflusst ist, vermochte die vorliegende Arbeit nicht aufzuklären.

Im Vergleich zum Level 2 lieferte ein Level 3 in Verbindung mit der Bearbeitung motivierender fahrfremder Tätigkeiten (FFT) ein besseres Bild in Bezug auf den Fahrerzustand. Statt zu ermüden, konnte die Müdigkeit über 25 Minuten in Level 3 unterdrückt werden – während dieser Zeit waren die Probanden sogar im Mittel wacher als zu Versuchsbeginn. Erst im Anschluss zeichnete sich eine Ermüdung der Fahrer ab. Ein sehr interessanter Zusammenhang wurde zwischen Ermüdung während Level 3 und dem Vertrauen in die Fahrfunktion nachgewiesen: Wenn Fahrer ein geringeres Vertrauen besaßen, führten sie mehr Kontrollblicke aus. Im Prinzip gehen sie derweil kurzen Überwachungssequenzen nach – ihre Aufmerksamkeit ist nicht mehr vollständig auf die FFT gerichtet. Dieses Verhalten ging mit einer höheren Müdigkeitsentwicklung einher als dies bei Teilnehmern zu beobachten war, die sich permanent mit ihrer FFT beschäftigten und der Regelungstätigkeit des Level 3 vertrauten.

Nicht nur die Level 3-Fahrt mit FFT war in der Lage, Ermüdungserscheinungen effektiv zu verhindern. Auch Übernahmeaufforderungen zum manuellen Fahren nach 28 Minuten profilieren sich sowohl nach Level 3 als auch nach Level 2, einen Aktivierungseffekt herbeiführen zu können. In seiner Stärke war dieser Effekt vergleichbar mit dem Aktivierungspotential, welches durch das Erleben eines Automatisierungsfehlers während Level 2 hervorgerufen wurde. Ob die Aktivierung einer solchen Transition nachhaltiger ist als eine Fehlerpräsentation (also länger als zwei Minuten andauert) konnte vorliegende Arbeit nicht ermitteln. Es zeigte sich

jedoch, dass die Fahrer im Mittel vier Sekunden nach Level 2 benötigten, um die Fahrzeugsteuerung zu übernehmen. Da bei Level 2 faktisch eine unmittelbare Übernahmebereitschaft vorliegen muss, sind solche lange Zeiten aufgrund eines ermüdeten Zustands im realen Straßenverkehr sehr kritisch zu beurteilen. Level 3 kann hier per se aufgrund der Funktionsauslegung Abhilfe schaffen – aber auch hier müsste die Regelungstätigkeit durch die Funktion mindestens acht Sekunden aufrechterhalten werden, damit die FFT vor Übernahme beiseitegelegt werden können. Dadurch ist noch nicht gesagt, dass die Fahrer in dieser Zeit auch ein adäquates Situationsbewusstsein hergestellt haben. Hierauf gibt noch keine gegenwärtig vorliegende Studie eine zufriedenstellende Antwort, weshalb weitere, innovative Forschungsansätze dringend benötigt werden.

Sowohl Experiment 1 als auch Experiment 2 lieferte seinen Beitrag zur Abschätzung des Fahrerzustands in Level 2. Es konnten sowohl Hinweise auf eine Über- als auch eine Unterbeanspruchung durch Nutzung einer solchen Funktion gefunden werden, die typischerweise bei Bearbeitung von Vigilanzaufgaben vorherrschen. Ein zu niedriges Arousal-Niveau der „ermüdeten“ Probanden im ersten Experiment deutete auf eine Unterbeanspruchung hin, ein zu hoher Workload verbunden mit einem höheren Stresserleben spricht in Experiment zwei sicherlich für eine Überbeanspruchung. Prototypisch konnte damit der auf Basis einschlägiger Forschungsliteratur herausgearbeitete Dualismus zwischen Unter- und Überbeanspruchung bei Vigilanzaufgaben im Anwendungskontext greifbar gemacht werden. Ferner wurde bestätigt, dass sowohl aufgabenbezogene als auch schlafbezogene Müdigkeit bei Verwendung eines Level 2 auftreten kann.

Letztlich ist auch eine Level 3-Fahrt – wie bereits in verschiedenen Studien aufgezeigt wurde – kein Garant für einen optimalen Fahrerzustand: Wenn keine motivierenden FFT bearbeitet werden, oder diese häufig unterbrochen werden, begibt sich der Fahrer in einen Zustand, aus dem heraus sich sein Aktivierungsniveau verschlechtern kann. Es besteht nämlich die Gefahr, dass der Fahrer im hohen Maße eine „Überwacherrolle“ einnimmt. Dies sollte verhindert werden. Es wird sogar für weitere Forschungsvorhaben die Hypothese aufgestellt, dass ein Level 3 ohne FFT zu nutzen mit vergleichbaren fahrerzustandsbezogenen Nachteilen verbunden ist, wie in einem Level 2 mit permanenter Überwachungsnotwendigkeit zu fahren.

Wie abermals zu sehen ist, stellt automatisiertes Fahren besondere Anforderungen an den Fahrer und seinen Zustand. Daher sollte abschließend explizit auf den kritischen Fahrerzustand hingewiesen werden, der sich nach längerer Level 2-Nutzung in Form erhöhter Müdigkeit abzeichnen kann. Beide Experimente haben dies unabhängig voneinander gezeigt. Gelegentliche,

im besten Fall unregelmäßige Automatisierungsfehler, sowie Transitionen zum manuellen Fahren können hier Abhilfe leisten und (zumindest kurzzeitig) eine Aktivierung des Fahrers herbeiführen. Um dieses Problem aber umfassend zu umgehen, wäre eine Level 3-Funktion zu bevorzugen. Allerdings nur dann, wenn die fortwährende Bearbeitung mit motivierenden FFT sichergestellt ist und die Fahrdauer an dem vorherrschenden Fahrerzustand ausgerichtet wird. Damit wird unweigerlich eine Fahrerzustandserkennung vonnöten sein, die im besten Falle auch Anhaltspunkte für das notwendige Situationsbewusstsein bei manueller Steuerungsübernahme liefert. In den nächsten Jahren sollte die fahrerseitige Routine im Umgang mit diesen Automatisierungsfunktionen steigen, weil sie die Fahrzeugflotte stärker durchdringen werden. Letztlich kann so auch das Funktionsvertrauen gestärkt und die Verkehrssicherheit erhöht werden.

Sicherlich wird auch eine technologische Weiterentwicklung zu beobachten sein, die eine höhere Zuverlässigkeit der Funktionen zum Ziel hat. Wenn damit fahrerseitige Interventionen seltener notwendig werden, und nur noch in komplexen Verkehrssituationen erforderlich sind, die die Funktion nicht mehr selbstständig lösen kann, wird dies kritisch beurteilt. Aufgrund seines schlechten Fahrerzustands, seiner mangelnden Routine als auch seines unzureichenden Situationsbewusstseins könnte der Mensch dieser Aufgabe dann nicht gewachsen sein. Vor diesem Hintergrund würden noch höhere Automatisierungsfunktionen, konkret ein Level 4 benötigt (vgl. Abschnitt 2.1), um den menschlichen Fahrer nicht in Aufgaben zu drängen, für die er gänzlich ungeeignet ist. Aber auch eine Level 4-Fahrt endet, sobald das Domänenende naht. Für die manuelle Weiterfahrt spielt dann der Fahrerzustand ebenfalls eine nicht zu unterschätzende Rolle. Insofern sollte auch die zukünftige Entwicklung automatisierter Fahrfunktion mit Forschungsaktivitäten zur Mensch-Maschine-Interaktion begleitet werden – damit die menschlichen Leistungsgrenzen stets ihre Berücksichtigung finden.

6 Literaturverzeichnis

- Aaronson, L. S., Teel, C. S., Cassmeyer, V., Neuberger, G. B., Pallikkathayil, L., Pierce, J., ... & Wingate, A. (1999). Defining and measuring fatigue. *Image: The Journal of Nursing Scholarship*, 31(1), 45-50.
- Acioğlu, A., & Erçelebi, E. (2016). Real time eye detection algorithm for PERCLOS calculation. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (pp. 1641-1644). IEEE.
- Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., & Randler, C. (2012). Circadian typology: a comprehensive review. *Chronobiology International*, 29(9), 1153-1175.
- Åkerstedt, T. (2000). Consensus statement: fatigue and accidents in transport operations. *Journal of Sleep Research*, 9(4), 395-395.
- Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2), 29-37.
- Andersen, G. J. (2011). Sensory and perceptual factors in the design of driving simulation displays. In D. L. Fisher, M. Rizzo, J. K. Caird & J. D. Lee (Eds.), *Handbook of Driving Simulation for Engineering, Medicine, and Psychology* (pp. 8-1-8-11). Boca Raton, FL: CRC Press.
- Anokhin, A., Steinlein, O., Fischer, C., Mao, Y., Vogt, P., Schalt, E., & Vogel, F. (1992). A genetic study of the human low-voltage electroencephalogram. *Human Genetics*, 90(1-2), 99-112.
- Antrobus, J. S., Singer, J. L., Goldstein, S., & Fortgang, M. (1970). Section of psychology: mindwandering and cognitive structure. *Transactions of the New York Academy of Sciences*, 32(2 Series II), 242-252.
- Aronson, E., Wilson, T. D., & Akert, R. M. (2008). *Sozialpsychologie*. Pearson Studium, 6. Auflage, Pearson Deutschland GmbH.
- Atchley, P., Chan, M., & Gregersen, S. (2014). A strategically timed verbal task improves performance and neurophysiological alertness during fatiguing drives. *Human Factors*, 56(3), 453-462.
- Bainbridge, L. (1983). Ironies of automation. In *Analysis, Design and Evaluation of Man-Machine Systems* (pp. 129-135). Pergamon.
- Baltodano, S., Sibi, S., Martelaro, N., Gowda, N., & Ju, W. (2015). The RRADS platform: a real road autonomous driving simulator. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 281-288).

- Barr, L., Howarth, H., Popkin, S., & Carroll, R. J. (2005). A review and evaluation of emerging driver fatigue detection measures and technologies. *National Transportation Systems Center, Cambridge. US Department of Transportation, Washington.*
- Baum, L. F. (1900). *The Wonderful Wizard of Oz. Chicago: George M. Hill.*
- Bazanov, O. M., & Vernon, D. (2014). Interpreting EEG alpha activity. *Neuroscience & Biobehavioral Reviews*, 44, 94-110.
- Becker, A. B., Warm, J. S., Dember, W. N., & Hancock, P. A. (1991). Effects of feedback on perceived workload in vigilance performance. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 35, No. 20, pp. 1491-1494). Sage CA: Los Angeles, CA: SAGE Publications.
- Beggiato, M. (2015). *Changes in motivational and higher-level cognitive processes when interacting with in-vehicle automation* (Doctoral Dissertation, University of Technology Chemnitz).
- Bengler, K., Omozik, K., & Müller, A. I. (2019). The Renaissance of Wizard of Oz (WoOz) – Using the WoOz methodology to prototype automated vehicles. *Proceedings of the Human Factors and Ergonomics Society Europe.*
- Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1), 527-570.
- Berghöfer, F. L., Purucker, C., Naujoks, F., Wiedemann, K., & Marberger, C. (2018). Prediction of take-over time demand in conditionally automated driving – results of a real world driving study. In *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2018 Annual Conference* (pp. 69-81).
- Berry, P. (2017). The stomach—both a target and potential contributor in space sickness. *Nausea and Vomiting*, 43-58.
- Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., & Strayer, D. L. (2018). 80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 1878-1882). Sage CA: Los Angeles, CA: SAGE Publications.
- Birbaumer, N., & Schmidt, R. T. (2010). *Biologische Psychologie. 7., überarb. und erg. Aufl. Heidelberg.*
- Braby, C. D., Harris, D., & Muir, H. C. (1993). A psychophysiological approach to the assessment of work underload. *Ergonomics*, 36(9), 1035-1042.
- Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon.
- Brookhuis, K. A., De Waard, D., & Janssen, W. H. (2019). Behavioural impacts of advanced driver assistance systems – an overview. *European Journal of Transport and Infrastructure Research*, 1(3).

- Brown, I. D. (1995). Methodological issues in driver fatigue research. In L. Hartley (Hrsg.), *Fatigue and Driving: Driver Impairment, Driver Fatigue and Driving Simulation* (pp. 155-166). London: Taylor & Francis.
- Bubb, H., & Bengler, K. (2015). Fahrerassistenz. In Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (Eds.), *Automobilergonomie* (pp. 525-582), Springer Vieweg, Wiesbaden.
- Bundesgesetzblatt Teil I Nr. 38. (2017). Achtes Gesetz zur Änderung des Straßenverkehrsgesetzes. Bonn: Bundesanzeiger Verlag.
- Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) (2015). *Strategie automatisiertes und vernetztes Fahren*. Berlin: Bundesministerium für Verkehr und digitale Infrastruktur (BMVI).
- Carlson, N. R. (2004). *Physiologische Psychologie* (Vol. 8). München: Pearson Studium.
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in Experimental Social Psychology* (Vol. 24, pp. 201-234). Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Comstock, J. R., Harris, R. L., & Pope, A. T. (1988). Physiological assessment of task underload. *Space Operations Automation and Robotics (SOAR '88) Workshop*, Dayton, Ohio.
- Craig, A. (1987). Signal detection theory and probability matching apply to vigilance. *Human Factors*, 29(6), 645-652.
- Cunningham, S., Scerbo, M. W., & Freeman, F. G. (2000). The electrocortical correlates of daydreaming during vigilance tasks. *Journal of Mental Imagery*.
- Czeisler, C. A., Duffy, J. F., Shanahan, T. L., Brown, E. N., Mitchell, J. F., Rimmer, D. W., ... & Dijk, D. J. (1999). Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*, 284(5423), 2177-2181.
- Darshana, S., Fernando, D., Jayawardena, S., Wickramanayake, S., & DeSilva, C. (2014). Efficient PERCLOS and gaze measurement methodologies to estimate driver attention in real time. In *2014 5th International Conference on Intelligent Systems, Modelling and Simulation* (pp. 289-294). IEEE.
- Davies, D. R., & Parasuraman, R. (1982). *The Psychology of Vigilance*. Academic Press.
- De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27, 196-217.

- Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 9(4), 673-690.
- Desmond, P. A., & Hancock, P. A. (2001). Active and Passive Fatigue States. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 455-465). Mahwah, NJ: L. Erlbaum.
- Deutscher Verkehrssicherheitsrat (DVR) (2012). *Vision Zero – Grundlagen und Strategien*. (Schriftenreihe Verkehrssicherheit Volume 16). Bonn: Deutscher Verkehrssicherheitsrat (DVR).
- Di Stasi, L. L., Renner, R., Catena, A., Cañas, J. J., Velichkovsky, B. M., & Pannasch, S. (2012). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data. *Transportation Research Part C: Emerging Technologies*, 21(1), 122-133.
- Dierkes, F., Friedrich, B., Heinrich, T., Hoffmann, S., Maurer, M., Reschka, A., ... & Vogt, S. (2019). Infrastrukturbedarf automatisierten Fahrens – Grundlagenprojekt. *Berichte der Bundesanstalt für Straßenwesen, Unterreihe Fahrzeugtechnik, (F130)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- Dijk, D. J., Groeger, J. A., Stanley, N., & Deacon, S. (2010). Age-related reduction in daytime sleep propensity and nocturnal slow wave sleep. *Sleep*, 33(2), 211-223.
- Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, 4, 4-14.
- Dinges, D. F., Mallis, M. M., Maislin, G., & Powell, J. W. (1998). *Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management* (No. DOT-HS-808-762). United States. National Highway Traffic Safety Administration.
- Dinges, D. F., Orne, M. T., Whitehouse, W. G., & Orne, E. C. (1987). Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. *Sleep*, 10(4), 313-329.
- Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, 17(6), 652-655.
- Dorrian, J., Rogers, N. L., & Dinges, D. F. (2004). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In *Sleep Deprivation* (pp. 67-98). CRC Press.
- Drummond, S. P., Bischoff-Grethe, A., Dinges, D. F., Ayalon, L., Mednick, S. C., & Meloy, M. J. (2005). The neural basis of the psychomotor vigilance task. *Sleep*, 28(9), 1059-1068.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64.

- Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., & Uresin, Y. (2004). Alpha rhythm of the EEG modulates visual detection performance in humans. *Cognitive Brain Research*, 20(3), 376-383.
- Fairclough, S. H. (2001). Mental effort regulation and the functional impairment of the driver. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 479-502). Mahwah, NJ: L. Erlbaum.
- Farmer, R., & Sundberg, N. D. (1986). Boredom proneness – the development and correlates of a new scale. *Journal of Personality Assessment*, 50(1), 4-17.
- Feldhütter, A., Feierle, A., Kalb, L., & Bengler, K. (2018). A new approach for a real-time non-invasive fatigue assessment system for automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 1669-1673). Sage CA: Los Angeles, CA: SAGE Publications.
- Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2017). How the duration of automated driving influences take-over performance and gaze behavior. In *Advances in Ergonomic Design of Systems, Products and Processes* (pp. 309-318). Springer, Berlin, Heidelberg.
- Feldhütter, A., Härtwig, N., Kurpiers, C., Hernandez, J. M., & Bengler, K. (2018a). Effect on mode awareness when changing from conditionally to partially automated driving. In *Congress of the International Ergonomics Association* (pp. 314-324). Springer, Cham.
- Feldhütter, A., Hecht, T., & Bengler, K. (2018b). Fahrerspezifische Aspekte beim hochautomatisierten Fahren. *FAT Schriftenreihe*, (307). Berlin: Forschungsvereinigung Automobiltechnik e.V. (FAT).
- Fisk, A. D., & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. *Human Factors*, 23(6), 737-750.
- Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., DeLongis, A., & Gruen, R. J. (1986). Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, 50(5), 992.
- Fors, C., Ahlstrom, C., & Anund, A. (2018). A comparison of driver sleepiness in the simulator and on the real road. *Journal of Transportation Safety & Security*, 10(1-2), 72-87.
- Forsman, P. M., Vila, B. J., Short, R. A., Mott, C. G., & Van Dongen, H. P. (2013). Efficient driver drowsiness detection at moderate levels of drowsiness. *Accident Analysis & Prevention*, 50, 341-350.
- Freeman, F. G., Mikulka, P. J., Scerbo, M. W., & Scott, L. (2004). An evaluation of an adaptive automation system using a cognitive vigilance task. *Biological Psychology*, 67(3), 283-297.
- Frey, A. (2016). Statischer und dynamischer Fahrsimulator im Vergleich-Wahrnehmung von Abstand und Geschwindigkeit. *Berichte der Bundesanstalt für Straßenwesen, Unterreihe Fahrzeugtechnik, (F115)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).

- Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., ... & Lotz, C. (2012). Rechtsfolgen zunehmender Fahrzeugautomatisierung. *Berichte der Bundesanstalt für Straßenwesen, Unterreihe Fahrzeugtechnik, (F83)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- Gasser, T. M., Frey, A. T., Seeck, A., & Auerswald, R. (2017). Comprehensive definitions for automated driving and ADAS. In *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)* National Highway Traffic Safety Administration.
- Gasser, T., Schmidt, E. A., Bengler, K., Chiellino, U., Diederichs, F., Eckstein, L., ... & Hoyer, R. (2015). Bericht zum Forschungsbedarf. Runder Tisch Automatisiertes Fahren – AG Forschung. *Bundesanstalt für Straßenwesen, Germany*.
- Gasser, T. M., & Westhoff, D. (2012). BASt-study: Definitions of automation and legal issues in Germany. In *Proceedings of the 2012 Road Vehicle Automation Workshop*. Automation Workshop.
- Geiwitz, P. J. (1966). Structure of boredom. *Journal of Personality and Social Psychology*, 3(5), 592.
- Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., & Stern, R. M. (2001). A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine*, 72(2), 115.
- Gimeno, P. T., Cerezuela, G. P., & Montañés, M. C. (2006). On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures. *International Journal of Vehicle Design*, 42(1/2), 67-86.
- Gold, C. G. (2016). *Modeling of take-over performance in highly automated vehicle guidance* (Dissertation, Technische Universität München).
- Goldenbeld, C., & Nikolaou, D. (2019). ESRA2 (E-Survey of Road users' Attitudes) Thematic report No. 4: Driver fatigue.
- Golding, J. F. (2006). Predicting individual differences in motion sickness susceptibility by questionnaire. *Personality and Individual Differences*, 41(2), 237-248.
- Goldman, R. I., Stern, J. M., Engel Jr, J., & Cohen, M. S. (2002). Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport*, 13(18), 2487.
- Gonçalves, J., Happee, R., & Bengler, K. (2016). Drowsiness in conditional automation: proneness, diagnosis and driving performance effects. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 873-878). IEEE.
- Gordon, A., Wilkinson, R., McGown, A., & Jovanoska, S. (1997). The psychometric properties of the Boredom Proneness Scale: An examination of its validity. *Psychological Studies*.

- Grace, R., Byrne, V. E., Bierman, D. M., Legrand, J. M., Gricourt, D., Davis, B. K., ... & Carnahan, B. (1998). A drowsy driver detection system for heavy vehicles. In *17th DASC. AIAA/IEEE/SAE. Digital Avionics Systems Conference. Proceedings (Cat. No. 98CH36267)* (Vol. 2, pp. I36-1). IEEE.
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Vol. 1). New York: Wiley.
- Griefahn, B., Künemund, C., Bröde, P., & Mehnert, P. (2001). Zur Validität der deutschen Übersetzung des Morningness-Eveningness-Questionnaires von Horne und Östberg: The Validity of a German Version of the Morningness-Eveningness-Questionnaire Developed by Horne and Östberg. *Somnologie*, 5(2), 71-80.
- Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*, 45(3), 349-359.
- Gruschwitz, D., Hölscher, J., Raudszus, D., & Schulz, A. (2020). Marktdurchdringung von Fahrzeugsicherheitssystemen in Pkw 2017. *Berichte der Bundesanstalt für Straßenwesen, Unterreihe Mensch und Sicherheit, (M295)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- Gunn, D. V., Warm, J. S., Nelson, W. T., Bolia, R. S., Schumsky, D. A., & Corcoran, K. J. (2005). Target acquisition with UAVs: Vigilance displays and advanced cuing interfaces. *Human Factors*, 47(3), 488-497.
- Guralnick, M. J., & Harvey, K. G. (1970). Response requirements and performance in a visual vigilance task. *Psychonomic Science*, 20(4), 215-217.
- Hargutt, V. (2007). „Müdigkeit im Verkehr – Entstehung, Unfallrisiko und Gegenmaßnahmen“. Abgerufen von: <https://www.yumpu.com/de/document/read/19456323/interdisziplinares-zentrum-fur-verkehrswissenschaften-deine-haut> [24.02.2020].
- Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.
- Hartmann, E. (1973). *The Functions of Sleep* (No. 13). Yale University Press.
- Head, H. (1923). The Conception of Nervous and Mental Energy.(II.)" Vigilance; A Physiological State of the Nervous System.". *British Journal of Psychology*.
- Hecht, T., Feldhütter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., & Bengler, K. (2018). A Review of Driver State Monitoring Systems in the Context of Automated Driving. In *Congress of the International Ergonomics Association* (pp. 398-408). Springer, Cham.
- Helton, W. S. (2004). Validation of a short stress state questionnaire. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 11, pp. 1238-1242). Sage CA: Los Angeles, CA: SAGE Publications.

- Helton, W. S., & Russell, P. N. (2012). Brief mental breaks and content-free cues may not keep you focused. *Experimental Brain Research*, 219(1), 37-46.
- Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human Factors*, 58(3), 509-519.
- Hill, A. B., & Perkins, R. E. (1985). Towards a model of boredom. *British Journal of Psychology*, 76(2), 235-240.
- Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., ... & Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science*, 4(1-2), 89-112.
- Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45(1-3), 73-93.
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology*, 10(4), 431-436.
- Hoeger, R., Zeng, H., Hoess, A., Kranz, T., Boverie, S., Strauss, M., . . . & Stratil, H. (2011). Final report, deliverable D61. 1. *Highly Automated Vehicles for Intelligent Transport (HAVEit). 7th Framework Programme*. (EU).
- Hohm, A., Klejnowski, L., Skibinski, S., Bengler, K., Berger, S., Vetter, J., . . . & Stürmer, T. (2018). *KO-HAF - Kooperatives Hochautomatisiertes Fahren: (Projektübergreifender Schlussbericht)*. Abgerufen von: https://www.ko-haf.de/fileadmin/user_upload/projekt/19S14002_Ko-HAF_partner%C3%BCbergreifender-Schlussbericht_final.pdf [09.04.2020].
- Holte, H. (2018). Bedarfsanalysen im Kontext von Lebenslagen, Lebensstilen und verkehrssicherheitsrelevanten Erwartungen. *Berichte der Bundesanstalt für Straßenwesen, Unterreihe Mensch und Sicherheit, (M285)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*.
- Howell, W. C., Johnston, W. A., & Goldstein, I. L. (1966). Complex monitoring and its relation to the classical problem of vigilance. *Organizational Behavior and Human Performance*, 1(2), 129-150.
- Huemer, A. K., & Vollrath, M. (2012). Ablenkung durch fahrfremde Tätigkeiten – Machbarkeitsstudie. *Berichte der Bundesanstalt für Straßenwesen, Unterreihe Mensch und Sicherheit, (M225)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- IBM SPSS (2012). IBM SPSS statistics version 21. *Boston, Mass: International Business Machines Corp*, 126.

- ISO 14198. (2012). *Road Vehicles - Ergonomic Aspects of Transport Information and Control Systems - Calibration Tasks for Methods which Assess Driver Demand Due to the Use of In-vehicle Systems*. ISO/TS 14198:2012(en).
- ISO 26262. (2018). *Road Vehicles – Functional Safety*. ISO/TS 26262:2018(en).
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513-541.
- Jankowski, K. S., & Ciarkowska, W. (2008). Diurnal variation in energetic arousal, tense arousal, and hedonic tone in extreme morning and evening types. *Chronobiology International*, 25(4), 577-595.
- Jarosch, O., Gold, C., Naujoks, F., Wandtner, B., Marberger, C., Weidl, G., & Schrauf, M. (2019). The impact of Non-Driving Related Tasks on Take-over Performance in Conditionally Automated Driving – A Review of the Empirical Evidence. In 9. *Tagung Automatisiertes Fahren*.
- Jarosch, O., Kuhnt, M., Paradies, S., & Bengler, K. (2017). It's out of our hands now! Effects of non-driving related tasks during highly automated driving on drivers' fatigue. In: *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 26-29, 2017, Manchester Village, Vermont*. Iowa City, IA: Public Policy Center, University of Iowa, 2017: 319-325.
- Jarosch, O., Paradies, S., Feiner, D., & Bengler, K. (2019a). Effects of non-driving related tasks in prolonged conditional automated driving – A Wizard of Oz on-road approach in real traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65, 292-305.
- Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53-71.
- Johansson, E., Engström, J., Cherri, C., Nodari, E., Toffetti, A., Schindhelm, R., & Gelau, C. (2004). Review of existing techniques and metrics for IVIS and ADAS assessment. *Adaptive Integrated Driver Vehicle Interface (AIDE) Product number: IST-1-507674-IP*.
- Johns, M. (1998). Rethinking the assessment of sleepiness. *Sleep Medicine Reviews*, 2(1), 3-15.
- Johnston, W. A., Howell, W. C., & Williges, R. C. (1969). The components of complex monitoring. *Organizational Behavior and Human Performance*, 4(2), 112-124.
- Joos (2018). Fünfmal Gähnen, dreimal Kratzen – Bin ich müde? Auftreten und Zusammenhänge von Müdigkeitsindikatoren im Verlauf von Fahrermüdigkeit (Bachelorarbeit). *Technische Universität Braunschweig, Lehrstuhl für Ingenieur- und Verkehrspsychologie, Braunschweig, Germany*.

- Josten, J., Zlocki, A., & Eckstein, L. (2016). Untersuchung der Bewältigungsleistung des Fahrers von kurzfristig auftretenden Wiederübernahmesituationen nach teilautomatischem, freihändigen Fahren. *FAT Schriftenreihe*, (289). Berlin: Forschungsvereinigung Automobiltechnik e.V. (FAT).
- Katz, J. W. (2011). Method for relieving motion sickness and related apparatus. *U.S. Patent Application Nr. 12/750,358*.
- Kecklund, G., & Åkerstedt, T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics*, 36(9), 1007-1017.
- Kerkhof, G. A., & Van Dongen, H. P. (1996). Morning-type and evening-type individuals differ in the phase position of their endogenous circadian oscillator. *Neuroscience Letters*, 218(3), 153-156.
- Klamroth, A., Zerbe, A., Marx, T. (2019). Transitionen bei Level 3-Automation: Einfluss der Verkehrsumgebung auf die Bewältigungsleistung des Fahrers während Realfahrten. *FAT Schriftenreihe*, (323). Berlin: Forschungsvereinigung Automobiltechnik e.V. (FAT).
- Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol. Suppl.*, 52(3), 3-6.
- Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews*, 53(1), 63-88.
- Kolrep, H., Rimini-Döring, M., Oehme, A., & Jürgensohn, T. (2005). Wie sieht müde aus? – Entwicklung und Validierung einer Skala zur Müdigkeitsbewertung von Kraftfahrern. 6. *Berliner Werkstatt Mensch-Maschine Systeme „Zustandserkennung und Systemgestaltung“*, S. 12.
- Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, 66, 18-31.
- Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. In *Proceedings of the XV International Conference on Human Computer Interaction* (pp. 1-7).
- Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, 3, 2403-2409.
- Kreuzmair, C., Gold, C., & Meyer, M. L. (2017). The influence of driver fatigue on take-over performance in highly automated vehicles. In *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration* (pp. 1-7).

- Kundinger, T., Riener, A., Sofra, N., & Weigl, K. (2018). Drowsiness Detection and Warning in Manual and Automated Driving: Results from Subjective Evaluation. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 229-236).
- Kundinger, T., Riener, A., Sofra, N., & Weigl, K. (2020). Driver drowsiness in automated and manual driving: insights from a test track study. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (pp. 369-379).
- Lal, S. K., & Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, 55(3), 173-194.
- Lamond, N., Dawson, D. R. E. W., & Roach, G. D. (2005). Fatigue assessment in the field: validation of a hand-held electronic psychomotor vigilance task. *Aviation, Space, and Environmental Medicine*, 76(5), 486-489.
- Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, 1(1), 6-21.
- Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: further investigations of sustained attention to response. *Neuropsychologia*, 37(6), 661-670.
- Manstetten, D., Marberger, C., & Beruscha, F. (2019). Wizard-of-Oz Experiments in Real Traffic – Can they Restart Human Factors?. *Hands off, Human Factors off? Welche Rolle spielen Human Factors in der Fahrzeugautomation?*, 21.
- Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and applying the concept of “driver availability” in automated driving. In *International Conference on Applied Human Factors and Ergonomics* (pp. 595-605). Springer, Cham.
- Masters, C. L., Simms, G., Weinman, N. A., Multhaup, G., McDonald, B. L., & Beyreuther, K. (1985). Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proceedings of the National Academy of Sciences*, 82(12), 4245-4249.
- Matthews, G., & Desmond, P. A. (1998). Personality and multiple dimensions of task-induced fatigue: A study of simulated driving. *Personality and Individual Differences*, 25(3), 443-458.
- Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology: Section A*, 55(2), 659-686.
- Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three. *Personality Psychology in Europe*, 7, 335-350.
- May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(3), 218-224.

- McKinley, R. A., McIntire, L. K., Schmidt, R., Pinchak, A., Caldwell, J. L., Repperger, D. W., & Kane, M. (2010). *Evaluation of Eye Metrics as a Detector of Fatigue* (No. AFRL-RH-WP-JA-2010-0002). Air Force Research Lab Wright-Patterson AFB OH Human Effectiveness Directorate.
- Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45(3), 594-597.
- Merrifield, C. (2010). Characterizing the Psychophysiological Signature of Boredom (Master Thesis). *University of Waterloo, Ontario, Canada*.
- Metz, B., Wörle, J., Hanig, M., Schmitt, M., & Lutz, A. (2020). Repeated Usage of an L3 Motorway Chauffeur: Change of Evaluation and Usage. *Information*, 11(2), 114.
- Miglani, A., Diels, C., & Terken, J. (2016). Compatibility between trust and non-driving related tasks in UI design for highly and fully automated driving. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 75-80).
- Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction becomes engagement in automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 1676-1680). Sage CA: Los Angeles, CA: SAGE Publications.
- Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, 38(2), 311-322.
- Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, 27, 84-92.
- Mouloua, M., Deaton, J., & Hitt, J. M. (2001). Automation and Workload in Aviation Systems. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 334-350). Mahwah, NJ: L. Erlbaum.
- Müller, A. I., Weinbeer, V., & Bengler, K. (2019). Using the wizard of Oz paradigm to prototype automated vehicles: methodological challenges. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 181-186).
- Muscio, B. (1921). Is a Fatigue Test Possible?. *British Journal of Psychology*.
- National Highway Traffic & Safety Administration (NHTSA). (2017). *ODI Resume*. Abgerufen von <https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF> [18.04.2020].
- National Transportation Safety Board (NTSB). (2018). *Preliminary Report Highway HWY18FH011*. Abgerufen von <https://goo.gl/cDv2kf> [18.04.2020].
- National Transportation Safety Board (NTSB). (2018a). *Preliminary Report Highway HWY18MH010*. Abgerufen von <https://goo.gl/2C6ZCH> [18.04.2020].

- National Transportation Safety Board (NTSB). (2020). *Collision Between a Sport Utility Vehicle Operating With Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018, HWY18FH011*. Abgerufen von <https://www.nts.gov/news/events/Documents/2020-HWY18FH011-BMG-abstract.pdf> [18.04.2020].
- Naujoks, F., Höfling, S., Purucker, C., & Zeeb, K. (2018). From partial and high automation to manual driving: relationship between non-driving related tasks, drowsiness and take-over performance. *Accident Analysis & Prevention*, 121, 28-42.
- Nedergaard, M., & Goldman, S. A. (2016). Brain drain. *Scientific American*, 314(3), 44.
- Neubauer, C., Matthews, G., & Saxby, D. (2014). Fatigue in the automated vehicle: do games and conversation distract or energize the driver?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2053-2057). Sage CA: Los Angeles, CA: SAGE Publications.
- Nguyen, H. A. (2018). *Using Personality Traits and Chronotype for Personalized Feedback in a Sleep Web App* (Doctoral Dissertation, Worcester Polytechnic Institute).
- Niederl, T. (2008). *Untersuchungen zu kumulativen psychischen und physiologischen Effekten des fliegenden Personals auf der Kurzstrecke* (Dissertation, Universität Kassel).
- Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., ... & Sandberg, H. (2004). HMI and safety-related driver performance. *Haste Deliverable*.
- Otmani, S., Pebayle, T., Roge, J., & Muzet, A. (2005). Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & Behavior*, 84(5), 715-724.
- Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I., Bekiaris, E., & Maglaveras, N. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology*, 118(9), 1906-1922.
- Parasuraman, R., & Davies, D. R. (1976). Decision theory analysis of response latencies in vigilance. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 578.
- Parasuraman, R., & Mouloua, M. (1987). Interaction of signal discriminability and task type in vigilance decrement. *Perception & Psychophysics*, 41(1), 17-22.
- Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue?. *Physiology & Behavior*, 93 (1-2), 369-378.
- Phillips, R. O. (2015). A review of definitions of fatigue—And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour*, 29, 48-56.

- Pigeau, R. A., Angus, R. G., O'Neill, P., & Mack, I. (1995). Vigilance latencies to aircraft detection among NORAD surveillance operators. *Human Factors*, 37(3), 622-634.
- Pinel, J. P., Barnes, S. J., & Pauli, P. (2019). *Biopsychologie* (Vol. 10). München: Pearson Studium.
- Platho, C., Pietrek, A., & Kolrep, H. (2013). Erfassung der Fahrerermüdung. *Berichte der Bundesanstalt für Straßenwesen. Unterreihe Fahrzeugtechnik, (F89)*. Bergisch Gladbach: Bundesanstalt für Straßenwesen (BASt).
- Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40(1-2), 187-195.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3-25.
- Preckel, F., Lipnevich, A. A., Schneider, S., & Roberts, R. D. (2011). Chronotype, cognitive abilities, and academic achievement: A meta-analytic investigation. *Learning and Individual Differences*, 21(5), 483-492.
- Radlmayr, J., Feldhütter, A., Frey, A., Jarosch, O., Marberger, C., Naujoks, F., . . . & Bengler, K. (2018). Drowsiness and fatigue in conditionally automated driving – Towards an integrative framework. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting Europe Chapter, (HFES) 2018*.
- Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 2063-2067). Sage CA: Los Angeles, CA: Sage Publications.
- Ralph, B. C. W., Onderwater, K., Thomson, D. R., & Smilek, D. (2016). Disrupting monotony while increasing demand: benefits of rest and intervening tasks on vigilance. *Psychological Research*, 81, 432-444.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203-212.
- Reinprecht, K. (2013). *Der Einfluss von Müdigkeit auf das Fahr-Blick- & Sprachverhalten unter der Betrachtung verschiedener methodischer Ansätze und die Konsequenz für die Entwicklung von Müdigkeitsassistenzsystemen* (Dissertation, Technische Universität Braunschweig).
- Reyner, L. A., & Horne, J. A. (1998). Falling asleep whilst driving: are drivers aware of prior sleepiness? *International Journal of Legal Medicine*, 111(3), 120-123.
- Reyner, L. A., Wells, S. J., Mortlock, V., & Horne, J. A. (2012). 'Post-lunch' sleepiness during prolonged, monotonous driving—Effects of meal size. *Physiology & Behavior*, 105(4), 1088-1091.

- Roenneberg, T. (2012). What is chronotype? *Sleep and Biological Rhythms*, 10(2), 75-76.
- Rogé, J., Pebayle, T., & Muzet, A. (2001). Variations of the level of vigilance and of behavioural activities during simulated automobile driving. *Accident Analysis & Prevention*, 33(2), 181-186.
- SAE (2018). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems (Standard No. J3016)*. SAE International, June 2018.
- Sannita, W. G., Loizzo, A., Garbarino, S., Gesino, D., Massimilla, S., & Ogliastro, C. (1999). Adrenocorticotropin-related modulation of the human EEG and individual variability. *Neuroscience Letters*, 262(3), 147-150.
- Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied*, 19(4), 287.
- Scerbo, M. W. (1998). What's so boring about vigilance? In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 145-166). Washington, DC: APA Books.
- Scerbo, M. W. (2001). Stress, workload, and boredom in vigilance: A problem and an answer. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 267-278). Mahwah, NJ: L. Erlbaum.
- Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?. *Ergonomics*, 51(7), 982-1010.
- Schmidt, E. (2020). Effect of Thermal Stimuli on Passive Fatigue While Driving. In *Effects of Thermal Stimulation During Passive Driver Fatigue* (pp. 67-120). Springer Vieweg, Wiesbaden.
- Schmidt, E. A. (2010). *Die objektive Erfassung von Müdigkeit während monotoner Tagfahrten und deren verbale Selbsteinschätzung durch den Fahrer* (Dissertation, Heinrich-Heine-Universität Düsseldorf).
- Schmidt, E. A., Schrauf, M., Simon, M., Buchner, A., & Kincses, W. E. (2011). The short-term effect of verbally assessing drivers' state on vigilance indices during monotonous daytime driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(3), 251-260.
- Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, 41(5), 1087-1093.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1.

- Schömig, N., Befelein, D., Wiedemann, K., & Neukum, A. (2020). Methodische Aspekte und aktuelle inhaltliche Schwerpunkte bei der Konzeption experimenteller Studien zum hochautomatisierten Fahren. *FAT Schriftenreihe*, (324). Berlin: Forschungsvereinigung Automobiltechnik e.V. (FAT).
- Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., & Othersen, I. (2015). The interaction between highly automated driving and the development of drowsiness. *Procedia Manufacturing*, 3, 6652-6659.
- Schömig, N. & Kaussner, A. (2010). Evaluation of an Attention Monitor for drowsy driving. In A. Schieben & F. Flemisch (Eds.), *Validation of Concept on Optimum Task Repartition* (pp. 18-34). Public Deliverable of the HAVEit project D33.6.
- Siebert, F. W., Oehl, M., Höger, R., & Pfister, H. R. (2013). Discomfort in automated driving—the disco-scale. In *International Conference on Human-Computer Interaction* (pp. 337-341). Springer, Berlin, Heidelberg.
- Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., ... & Schrauf, M. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clinical Neurophysiology*, 122(6), 1168-1178.
- Sivak, M., & Schoettle, B. (2015). *Motion sickness in self-driving vehicles*. University of Michigan, Ann Arbor, Transportation Research Institute.
- Smit, A. S., Eling, P. A., Hopman, M. T., & Coenen, A. M. (2005). Mental and physical effort affect vigilance differently. *International Journal of Psychophysiology*, 57(3), 211-217.
- Smith, A. (2002). Effects of caffeine on human behavior. *Food and Chemical Toxicology*, 40(9), 1243-1255.
- Stanton, N. A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. *Safety Science*, 24(1), 35-49.
- Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2017). Driver behavior and workload in an on-road automated vehicle. In *Proceedings of the RSS2017 Conference*.
- Statistisches Bundesamt (Destatis). (2020). *Verkehr: Verkehrsunfälle 2018* (Fachserie 8 Reihe 7). Wiesbaden: Statistisches Bundesamt (Destatis).
- Steinlein, O., Anokhin, A., Yping, M., Schalt, E., & Vogel, F. (1992). Localization of a gene for the human low-voltage EEG on 20q and genetic heterogeneity. *Genomics*, 12(1), 69-73.
- Stern, J. A., Boyer, D., & Schroeder, D. J. (1994). *Blink rate as a measure of fatigue: A review*. Washington Univ St Louis Mo.
- Strelow (2018). *Was passiert, wenn man müde Auto fährt? – Eine Fahrstudie zu Müdigkeit und Fahrverhalten* (Bachelorarbeit, Technische Universität Braunschweig, Lehrstuhl für Ingenieur- und Verkehrspsychologie, Braunschweig, Germany).

- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4(4), 525-547.
- Telpaz, A., Rhindress, B., Zelman, I., & Tsimhoni, O. (2015). Haptic seat for automated driving: preparing the driver to take control effectively. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 23-30).
- Thomson, D. R., Besner, D., & Smilek, D. (2016). A critical examination of the evidence for sensitivity loss in modern vigilance tasks. *Psychological Review*, 123(1), 70.
- Tietze, H. (2001). Stages of fatigue during long duration driving reflected in alpha related events in the EEG. In *International Conference on Traffic and Transport Psychology-ICTTP 2000, held 4-7 September 2000, Berne, Switzerland*.
- Torsvall, L., & Åkerstedt, T. (1987). Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology*, 66(6), 502-511.
- Trutschel, U., Sirois, B., Sommer, D., Golz, M., & Edwards, D. (2011). PERCLOS: An Alertness Measure of the Past. In *Driving Assessment 2011: 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle*. University of Iowa, Iowa City.
- Tudor, M., Tudor, L., & Tudor, K. I. (2005). Hans Berger (1873-1941) – The History of Electroencephalography. *Acta Medica Croatica: Casopis Hrvatske Akademije Medicinskih Znanosti*, 59(4), 307.
- Ulmer, B. (2004). ADASE Final Report. *Advanced Driver Assistance Systems in Europe (ADASE) Contract number: IST-2000-28010*.
- Van Dijk, H., Schoffelen, J. M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *Journal of Neuroscience*, 28(8), 1816-1823.
- Vogelpohl, T., Vollrath, M., & Kühn, M. (2017). Übergabe von hochautomatisiertem Fahren zu manueller Steuerung. Teil 2: Müdigkeit und lange Fahrdauer als Einflussfaktoren auf die Sicherheit nach einer Übernahmeaufforderung. *Forschungsbericht/Unfallforschung der Versicherer (GDV)*, (47). Berlin: GDV.
- Vogelpohl, T., Vollrath, M., Kühn, M., Hummel, T., & Gehlert, T. (2016). Übergabe von hochautomatisiertem Fahren zu manueller Steuerung. Teil 1: Review der Literatur und Studie zu Übernahmezeiten. *Forschungsbericht/Unfallforschung der Versicherer (GDV)*, (39). Berlin: GDV.
- Vollrath, M. (2015). Fahrerassistenz. In Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (Eds.), *Automobilergonomie* (pp. 663-684), Springer Vieweg, Wiesbaden.

- Vollrath, M. (2018). *MUED - Handbuch*. Technische Universität Braunschweig, Lehrstuhl für Ingenieur- und Verkehrspsychologie. Verfügbar unter: http://cgi-psych2.rz.tubs.de/iv/software/Manual_MUED.pdf
- Vollrath, M. (2018a). *MUED – Macht Beurteilung von Muedigkeit einfach! (Version 8.1) [Microsoft Windows]*. Braunschweig: Technische Universität Braunschweig, Lehrstuhl für Ingenieur- und Verkehrspsychologie. Verfügbar unter: <https://www.tu-braunschweig.de/psychologie/verkehrspsychologie/software>
- Vollrath, M., Schleicher, S., & Gelau, C. (2011). The influence of Cruise Control and Adaptive Cruise Control on driving behaviour – A driving simulator study. *Accident Analysis & Prevention*, 43(3), 1134-1139.
- Wada, T. (2016). Motion sickness in automated vehicles. In *Advanced Vehicle Control: Proceedings of the 13th International Symposium on Advanced Vehicle Control (AVEC'16)* (pp. 169-174).
- Walch, M., Mühl, K., Kraus, J., Stoll, T., Baumann, M., & Weber, M. (2017). From car-driver-handovers to cooperative interfaces: Visions for driver – vehicle interaction in automated driving. In *Automotive User Interfaces* (pp. 273-294). Springer, Cham.
- Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. *Automation and Human Performance: Theory and Applications*, 183-200.
- Warm, J. S., Matthews, G., & Finomore Jr, V. S. (2018). Vigilance, workload, and stress. In *Performance Under Stress* (pp. 131-158). CRC Press.
- Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, 50(3), 433-441.
- Webb, W. B. (1988). Theoretical Presentation: An Objective Behavioral Model of Sleep. *Sleep*, 11(5), 488-496.
- Weinbeer, V., Bill, J. S., Baur, C., & Bengler, K. (2018). Automated driving: Subjective assessment of different strategies to manage drowsiness. In *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2017 Annual Conference* (pp. 5-17).
- Weinbeer, V., Muhr, T., Bengler, K., Baur, C., Radlmayr, J., & Bill, J. (2017). Highly automated driving: How to get the driver drowsy and how does drowsiness influence various take-over-aspects?. In 8. *Tagung Fahrerassistenz*.
- Weinman, J., Petrie, K. J., Moss-Morris, R., & Horne, R. (1996). The illness perception questionnaire: a new method for assessing the cognitive representation of illness. *Psychology and Health*, 11(3), 431-445.
- Weiß, C. (2007). Koffein. *Ernährungsumschau*, 54(4), 210.
- Wen, P., & Li, Y. (2006). EEG human head modelling based on heterogeneous tissue conductivity. *Australasian Physics & Engineering Sciences in Medicine*, 29(3), 235.

- Wever, R. A. (2013). *The Circadian System of Man: Results of Experiments under Temporal Isolation*. Springer Science & Business Media.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449-455.
- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering Psychology and Human Performance* (4th eds). Psychology Press.
- Wiegand, D. M., McClafferty, J., McDonald, S. E., & Hanowski, R. J. (2009). Development and evaluation of a naturalistic observer rating of drowsiness protocol: Final report. *National Surface Transportation Safety Center for Excellence (NSTSC; VTTI)*.
- Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis & Prevention*, 26(5), 571-581.
- Winner, H., Hakuli, S., Lotz, F., & Singer, C. (Eds.). (2016). *Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort*. Springer International Publishing.
- Wittmann, M., Dinich, J., Mellow, M., & Roenneberg, T. (2006). Social jetlag: misalignment of biological and social time. *Chronobiology International*, 23(1-2), 497-509.
- Wörle, J. & Metz, B. (2020). Driving with an L3 - motorway chauffeur: How do drivers use their driving time? In D. de Waard, A. Toffetti, L. Pietrantonio, T. Franke, J-F. Petiot, C. Dumas, A. Botzer, L. Onnasch, I. Milleville, and F. Mars (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference*.
- Wright, K. (Jr.), Badia, P., Myers, B., & Plenzler, S. (1997). Combination of bright light and caffeine as a countermeasure for impaired alertness and performance during extended sleep deprivation. *Journal of Sleep Research*, 6(1), 26-35.
- Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiagarajan, M., ... & Takano, T. (2013). Sleep drives metabolite clearance from the adult brain. *Science*, 342(6156), 373-377.
- Xiong, Y., Gao, J., Yang, Y., Yu, X., & Huang, W. (2016). Classifying driving fatigue based on combined entropy measure using EEG signals. *Int. J. Control Autom*, 9(3), 329-338.
- Yan, J. J., Kuo, H. H., Lin, Y. F., & Liao, T. L. (2016). Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing. In *2016 International Symposium on Computer, Consumer and Control (IS3C)* (pp. 243-246). IEEE.
- Yanko, M. R., & Spalek, T. M. (2014). Driving with the wandering mind: the effect that mind-wandering has on driving performance. *Human Factors*, 56(2), 260-269.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459-482.

- Young, K. & Regan, M. (2007). Driver distraction: A review of the literature. In: I.J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J.D. Irwin (Eds.). *Distracted Driving* (pp. 379-405). Sydney, NSW: Australasian College of Road Safety.
- Zhang, B., De Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 64, 285-307.
- Zhao, C., Zheng, C., Zhao, M., & Liu, J. (2010). Physiological assessment of driving mental fatigue using wavelet packet energy and random forests. *American Journal of Biomedical Sciences*, 2(3), 262-274.
- Zulley, J., & Hajak, G. (2005). Grundlegendes Wissen über den Schlaf. *Verhaltenstherapie*, 15(4), 212-218.
- Zwyghuizen-Doorenbos, A., Roehrs, T. A., Lipschutz, L., Timms, V., & Roth, T. (1990). Effects of caffeine on alertness. *Psychopharmacology*, 100(1), 36-39.

7 Anhang

Anhang A. Fragebogen Experiment 1, Teil 1

Wir möchten Sie bitten, den folgenden Fragebogen auszufüllen. Bitte kreuzen Sie immer **genau ein Kästchen** an und **lassen Sie keine Frage aus**. Es gibt keine richtigen oder falschen Antworten.

WICHTIGE Information: Unter einem **teilautomatisierten System** versteht man ein System, das das **Beschleunigen und Bremsen** und das **Lenken** übernimmt. Dabei **muss** der Fahrer das System jedoch dauerhaft überwachen und jederzeit zur vollständigen Übernahme der Fahrzeugführung bereit sein.

Können Sie sich generell vorstellen, in Ihrem Alltag so ein teilautomatisiertes System zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 auf jeden Fall

Verwenden Sie in Ihrem Fahrzeug regelmäßig folgende Assistenzsysteme?

- | | | | |
|--------------------------------------|-----------------------------|-------------------------------|---|
| - vorausschauender Notbremsassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Tempomat | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Abstandsregler | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Spurwechselassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Spurhalteassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Parklenkassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Fernlichtassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Verkehrszeichenassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Müdigkeitswarner | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |
| - Nachtsichtassistent | <input type="checkbox"/> ja | <input type="checkbox"/> nein | <input type="checkbox"/> bin mir nicht sicher |

Bitte bewerten Sie Ihre aktuelle Schläfrigkeit auf der folgenden Skala:

wach	wenig			mittel			stark			sehr stark		
0	-	0	+	-	0	+	-	0	+	-	0	+

Der folgende Teil des Fragebogens soll erfassen, ob und in welchen Fortbewegungsmitteln oder Situationen Sie bereits **die Reisekrankheit** erlebt haben. Sich reisekrank zu fühlen bedeutet, sich bei dem passiven Mitfahren in einem Fahrzeug mulmig zu fühlen, Übelkeit, Erbrechen oder andere Symptome zu erleben.

Bitte beantworten Sie den Fragebogen zunächst bezogen auf Ihre **Kindheitserfahrungen (vor dem 12. Lebensjahr)**. Falls Sie die Transport- oder Unterhaltungsmittel nicht verwendet haben, kreuzen Sie bitte „nicht zutreffend“ in der ersten Spalte an.

	Nicht zutreffend – nie verwendet	Nie Reise-krankheit erlebt	Selten Reise-krankheit erlebt	Manchmal Reise-krankheit erlebt	Häufig Reise-krankheit erlebt
Auto					
Bus oder Reisebus					
Zug					
Flugzeug					
Kleine Boote					
Schiffe, z.B. Fähren					
Schaukeln auf Spielplätzen					
Karusselle auf Spielplätzen					
Achterbahnen					

Beantworten Sie den Fragebogen nun bitte bezogen auf Ihre **Erfahrungen in den letzten 10 Jahren**. Falls Sie die Transport oder Unterhaltungsmittel nicht verwendet haben, kreuzen Sie bitte „nicht zutreffend“ in der ersten Spalte an.

	Nicht zutreffend – nie verwendet	Nie Reise-krankheit erlebt	Selten Reise-krankheit erlebt	Manchmal Reise-krankheit erlebt	Häufig Reise-krankheit erlebt
Auto					
Bus oder Reisebus					
Zug					
Flugzeug					
Kleine Boote					
Schiffe, z.B. Fähren					
Schaukeln auf Spielplätzen					
Karusselle auf Spielplätzen					
Achterbahnen					

Inwieweit stimmen Sie den folgenden Aussagen zur Automation zu?

WICHTIGE Information: Bitte stellen Sie sich folgendes automatisiertes System vor:

Das System ist so ausgelegt, dass **Beschleunigen und Bremsen** und das **Lenken** des Fahrzeugs vom System übernommen werden. Ihre Aufgabe ist es, das System **ständig zu überwachen** und **jederzeit zur vollständigen Übernahme der Fahrtätigkeit** bereit zu sein. Bewerten Sie also die folgenden Fragen so auf das System bezogen, wie Sie es sich momentan vorstellen. **1** steht hierbei immer für „geringe Zustimmung“ und **7** für „starke Zustimmung“.

1. Die Automation ist irreführend.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

2. Die Automation verhält sich auf eine unvorhersehbare Weise.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

3. Ich misstrau den Aktionen, Absichten oder Konsequenzen der Automation.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

4. Ich bin vorsichtig gegenüber der Automation.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

5. Die Handlungen der Automation führen zu nachteiligen oder schädlichen Konsequenzen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

6. Ich traue mir zu, die Automation zu nutzen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

7. Die Automation ist sicher.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

8. Die Automation ist glaubwürdig.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

9. Ich kann mich auf die Automation verlassen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

10. Die Automation ist zuverlässig.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

11. Ich kann der Automation vertrauen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

12. Ich bin mit der Automation vertraut.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Inwieweit treffen die folgenden Aussagen auf Ihre aktuellen Gefühle zu?

1 steht hierbei immer für „trifft überhaupt nicht zu“ und **5** für „trifft voll und ganz zu“.

1. unzufrieden

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

2. aufmerksam

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

3. deprimiert

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

4. traurig

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

5. aktiviert

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

6. ungeduldig

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

7. genervt

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

8. ärgerlich

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

9. gereizt

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

10. missgelaunt

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Inwieweit treffen die folgenden Aussagen auf Ihre aktuellen Gedanken zu?

1 steht hierbei immer für „trifft überhaupt nicht zu“ und **5** für „trifft voll und ganz zu“.

1. Ich möchte die Aufgabe erfolgreich ausführen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

2. Ich bin motiviert, die Aufgabe auszuführen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

3. Ich reflektiere mich selbst.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

4. Ich denke über mich selbst nach.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

5. Ich bin mir meiner Fähigkeiten sicher.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

6. Ich bin besorgt, was andere Leute von mir denken.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

7. Ich mache mir Gedanken darüber, was ich für einen Eindruck mache.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

8. Ich erwarte, dass ich die Aufgabe kompetent ausübe.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

9. Insgesamt habe ich das Gefühl, die Kontrolle zu haben.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

10. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

11. Ich habe darüber nachgedacht, wie ich mich fühlen würde, wenn ich eine Rückmeldung bekäme, wie gut ich bei der Aufgabe war.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Kreuzen Sie bitte an, inwieweit die folgenden Aussagen auf Sie zutreffen. Bitte kreuzen Sie immer genau **eine** Option an und lassen Sie keine Aussage aus.

Ich...	Trifft über- haupt nicht zu	Trifft eher nicht zu	Weder noch	Trifft eher zu	Trifft voll und ganz zu
1. ...bin eher zurückhaltend, reserviert.					
2. ...schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.					
3. ...erledige Aufgaben gründlich.					
4. ...bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.					
5. ...habe eine aktive Vorstellungskraft, bin phantasievoll.					
6. ...gehe aus mir heraus, bin gesellig.					
7. ...kann mich kalt und distanziert verhalten.					
8. ...bin bequem, neige zur Faulheit.					
9. ...werde leicht nervös und unsicher.					
10. ...habe nur wenig künstlerisches Interesse.					
11. ...bin rücksichtsvoll zu anderen, einfühlsam.					

Anhang B. Fragebogen Experiment 1, Teil 2

Wir möchten Sie bitten, den folgenden Fragebogen auszufüllen. Bitte kreuzen Sie **immer genau ein Kästchen** an und lassen Sie keine Frage aus. Es gibt keine richtigen oder falschen Antworten.

Bitte geben Sie an, wie es Ihnen momentan geht:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

„Ich fühle mich gut.“

„Ich fühle mich furchtbar, so als ob ich mich übergeben müsste.“

Bitte bewerten Sie Ihre aktuelle Schläfrigkeit auf der folgenden Skala:

wach	wenig			mittel			stark			sehr stark		
0	-	0	+	-	0	+	-	0	+	-	0	+

Wie zufrieden sind Sie mit dem eben verwendeten System?

gar nicht zufrieden ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 sehr zufrieden

Können Sie sich vorstellen, dieses System in Zukunft in Ihrem Alltag zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 auf jeden Fall

Was müsste noch verbessert werden, damit Sie das System definitiv nutzen würden?

Bitte beantworten Sie die folgenden Fragen in Bezug auf die gerade abgeschlossene Aufgabe:

Geistige Anforderung

Wie hoch waren die geistigen Anforderungen (z.B. Denken, Entscheiden, Erinnern, Hinsehen, Suchen) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Körperliche Anforderung

Wie hoch waren die körperlichen Anforderungen (z.B. Ziehen, Drücken, Drehen, Steuern) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Zeitliche Anforderung

**Wie hoch war das Tempo, mit der die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten?
War die Aufgabe langsam und geruhsam oder schnell und hektisch?**

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Aufgabenerfüllung

Wie erfolgreich haben Sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt?

gut ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 schlecht

Anstrengung

Wie sehr mussten Sie sich anstrengen, um Ihre Leistung zu erreichen?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Frustration**Wie verunsichert, entmutigt, gereizt und verärgert fühlen Sie sich während der Aufgabe?**sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark**Inwieweit treffen die folgenden Aussagen auf Ihre Gefühle während der Aufgabe zu?****1** steht hierbei immer für „trifft überhaupt nicht zu“ und **5** für „trifft voll und ganz zu“.**1. unzufrieden**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**2. aufmerksam**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**3. deprimiert**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**4. traurig**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**5. aktiviert**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**6. ungeduldig**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**7. genervt**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**8. ärgerlich**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**9. gereizt**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5**10. missgelaunt**☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Inwieweit treffen die folgenden Aussagen auf Ihre Gedanken während der Aufgabe zu?

1 steht hierbei immer für „trifft überhaupt nicht zu“ und **5** für „trifft voll und ganz zu“.

1. Ich wollte die Aufgabe erfolgreich ausführen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

2. Ich war motiviert, die Aufgabe auszuführen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

3. Ich habe mich selbst reflektiert.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

4. Ich habe über mich selbst nachgedacht.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

5. Ich war mir meiner Fähigkeiten sicher.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

6. Ich war besorgt, was andere Leute von mir denken.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

7. Ich habe mir Gedanken darüber gemacht, was ich für einen Eindruck mache.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

8. Ich habe die Aufgabe kompetent ausgeübt.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

9. Insgesamt hatte ich das Gefühl, die Kontrolle zu haben.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

10. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

11. Ich habe darüber nachgedacht, wie ich mich fühlen würde, wenn ich eine Rückmeldung bekäme, wie gut ich bei der Aufgabe war.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Inwieweit stimmen Sie den folgenden Aussagen zum System zu?

Sie haben diese Fragen bereits **vor** dem Fahrversuch einmal beantwortet. Bitte beantworten Sie diese nun noch einmal in Bezug auf die **Erfahrungen**, die Sie gerade mit dem System gemacht haben.

1 steht hierbei immer für „geringe Zustimmung“ und **7** für „starke Zustimmung“.

1. Die Automation ist irreführend.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

2. Die Automation verhält sich auf eine unvorhersehbare Weise.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

3. Ich misstrau den Aktionen, Absichten oder Konsequenzen der Automation.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

4. Ich bin vorsichtig gegenüber der Automation.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

5. Die Handlungen der Automation führen zu nachteiligen oder schädlichen Konsequenzen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

6. Ich traue mir zu, die Automation zu nutzen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

7. Die Automation ist sicher.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

8. Die Automation ist glaubwürdig.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

9. Ich kann mich auf die Automation verlassen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

10. Die Automation ist zuverlässig.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

11. Ich kann der Automation vertrauen.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

12. Ich bin mit der Automation vertraut.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Inwieweit stimmen Sie den folgenden Aussagen zum System zu?

1 steht hierbei immer für „geringe Zustimmung“ und **7** für „starke Zustimmung“.

1. Ich habe mich während der Fahrt sicher gefühlt.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

2. Ich hatte immer die Kontrolle über das System.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

3. Das System erhöht die Sicherheit.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

4. Die Nutzung des Systems ist unangenehm.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

5. Ich hatte stets das Gefühl, rechtzeitig eingreifen zu können.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

6. Das System geht kein unnötiges Risiko ein.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

7. Das System ist für mich eine zusätzliche Belastung.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

8. Ich wusste häufig nicht, was das System als nächstes machen wird.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

9. Das System entlastet mich als Fahrer.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

10. Ich fühlte mich durch das System in Gefahr gebracht.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

Auf welche Art und Weise wurde Ihrer Einschätzung nach das teilautomatisierte Fahren in diesem Versuch technisch umgesetzt?

Wie viele Stunden haben Sie in der letzten Nacht geschlafen? _____ Stunden

Haben Sie heute Koffein konsumiert? ☐ ja ☐ nein

Vielen Dank für Ihre Teilnahme!

Anhang C. Fragebogen Experiment 2, Teil 1

Wir möchten Sie bitten, den folgenden Fragebogen auszufüllen. Bitte kreuzen Sie immer **genau ein Kästchen** an und **lassen Sie keine Frage aus**. Es gibt keine richtigen oder falschen Antworten.

Bitte bewerten Sie Ihre aktuelle Schläfrigkeit auf der folgenden Skala:

wach	wenig			mittel			stark			sehr stark		
0	-	0	+	-	0	+	-	0	+	-	0	+

Verwenden Sie in Ihrem Fahrzeug regelmäßig folgende Assistenzsysteme?

- vorausschauender Notbremsassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Tempomat	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Abstandsregler	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Spurwechselassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Spurhalteassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Parklenkassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Fernlichtassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Verkehrszeichenassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Müdigkeitswarner	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher
- Nachtsichtassistent	<input type="checkbox"/> ja	<input type="checkbox"/> nein	<input type="checkbox"/> bin mir nicht sicher

Information zu automatisiertem Fahren („mit Überwachung“):

Beim automatisierten Fahren „mit Überwachung“ werden vom automatisierten Fahrsystem das **Beschleunigen und Bremsen** und das **Lenken** übernommen. Dabei **muss** der Fahrer das System **dauerhaft überwachen** und jederzeit zur vollständigen Übernahme der Fahrzeugführung bereit sein.

Können Sie sich generell vorstellen, in Ihrem Alltag ein solches automatisiertes System zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 auf jeden Fall

Bitte stellen Sie sich vor, dass Sie gerade automatisiertes Fahren „mit Überwachung“ nutzen.

Wie ist Ihre Zustimmung zu folgenden Aussagen:

Aussage	geringe Zustimmung 1	2	3	4	5	6	starke Zustimmung 7
1. Die Automation ist irreführend.							
2. Die Automation verhält sich auf eine unvorhersehbare Weise.							
3. Ich misstraue den Aktionen, Absichten oder Konsequenzen der Automation.							
4. Ich bin vorsichtig gegenüber der Automation.							
5. Die Handlungen der Automation führen zu nachteiligen oder schädlichen Konsequenzen.							
6. Ich traue mir zu, die Automation zu nutzen.							
7. Die Automation ist sicher.							
8. Die Automation ist glaubwürdig.							
9. Ich kann mich auf die Automation verlassen.							
10. Die Automation ist zuverlässig.							
11. Ich kann der Automation vertrauen.							
12. Ich bin mit der Automation vertraut.							

Information zu automatisiertem Fahren („mit fahrfremden Tätigkeiten“):

Beim automatisierten Fahren „mit fahrfremden Tätigkeiten“ werden vom automatisierten Fahrsystem das **Beschleunigen und Bremsen** und das **Lenken** übernommen. Das System muss **nicht dauerhaft überwacht** werden, d.h. der Fahrer darf sich währenddessen mit verschiedenen **anderweitigen (fahrfremden) Tätigkeiten** beschäftigen. Sobald das System an einer **Leistungsgrenze** angelangt ist oder ein **technischer Defekt** auftritt, muss der Fahrer die Fahrzeugführung mit einem Zeitpuffer von wenigen Sekunden übernehmen.

Können Sie sich generell vorstellen, in Ihrem Alltag ein solches automatisiertes System zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 auf jeden Fall

Bitte stellen Sie sich vor, dass Sie gerade automatisiertes Fahren „mit fahrfremden Tätigkeiten“ nutzen.

Wie ist Ihre Zustimmung zu folgenden Aussagen:

Aussage	geringe Zustimmung 1	2	3	4	5	6	starke Zustimmung 7
1. Die Automation ist irreführend.							
2. Die Automation verhält sich auf eine unvorhersehbare Weise.							
3. Ich misstrauere den Aktionen, Absichten oder Konsequenzen der Automation.							
4. Ich bin vorsichtig gegenüber der Automation.							
5. Die Handlungen der Automation führen zu nachteiligen oder schädlichen Konsequenzen.							
6. Ich traue mir zu, die Automation zu nutzen.							
7. Die Automation ist sicher.							
8. Die Automation ist glaubwürdig.							
9. Ich kann mich auf die Automation verlassen.							
10. Die Automation ist zuverlässig.							
11. Ich kann der Automation vertrauen.							
12. Ich bin mit der Automation vertraut.							

Inwieweit treffen die folgenden Aussagen auf Ihre aktuellen Gefühle zu?

Gefühl	Trifft über- haupt nicht zu 1	2	3	4	Trifft voll und ganz zu 5
1. unzufrieden					
2. aufmerksam					
3. deprimiert					
4. traurig					
5. aktiviert					
6. ungeduldig					
7. genervt					
8. ärgerlich					
9. gereizt					
10. missgelaunt					

Inwieweit treffen die folgenden Aussagen auf Ihre aktuellen Gedanken zu?

Gedanken	Trifft über- haupt nicht zu 1	2	3	4	Trifft voll und ganz zu 5
1. Ich möchte die Aufgabe erfolgreich ausführen.					
2. Ich bin motiviert, die Aufgabe auszuführen.					
3. Ich reflektiere mich selbst.					
4. Ich denke über mich selbst nach.					
5. Ich bin mir meiner Fähigkeiten sicher.					

6. Ich bin besorgt, was andere Leute von mir denken.					
7. Ich mache mir Gedanken darüber, was ich für einen Eindruck mache.					
8. Ich erwarte, dass ich die Aufgabe kompetent ausübe.					
9. Insgesamt habe ich das Gefühl, die Kontrolle zu haben.					
10. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben.					
11. Ich habe darüber nachgedacht, wie ich mich fühlen würde, wenn ich eine Rückmeldung bekäme, wie gut ich bei der Aufgabe war.					

Kreuzen Sie bitte an, inwieweit die folgenden Aussagen auf Sie zutreffen. Bitte kreuzen Sie immer genau **eine** Option an und lassen Sie keine Aussage aus.

Ich...	Trifft überhaupt nicht zu	Trifft eher nicht zu	Weder noch	Trifft eher zu	Trifft voll und ganz zu
1. ...bin eher zurückhaltend, reserviert.					
2. ...schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.					
3. ...erledige Aufgaben gründlich.					
4. ...bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.					
5. ...habe eine aktive Vorstellungskraft, bin phantasievoll.					
6. ...gehe aus mir heraus, bin gesellig.					
7. ...kann mich kalt und distanziert verhalten.					
8. ...bin bequem, neige zur Faulheit.					
9. ...werde leicht nervös und unsicher.					
10. ...habe nur wenig künstlerisches Interesse.					
11. ...bin rücksichtsvoll zu anderen, einfühlsam.					

Anhang D. Fragebogen Experiment 2, Teil 2

Wir möchten Sie bitten, den folgenden Fragebogen auszufüllen. Bitte kreuzen Sie **immer genau ein Kästchen** an und lassen Sie keine Frage aus. Es gibt keine richtigen oder falschen Antworten.

Bitte bewerten Sie Ihre aktuelle Schläfrigkeit auf der folgenden Skala:

wach	wenig			mittel			stark			sehr stark		
0	-	0	+	-	0	+	-	0	+	-	0	+

Nachfolgend erhalten Sie noch einmal die Automatisierungs-Beschreibung („mit Überwachung“):

Information zu automatisiertem Fahren („mit Überwachung“):

Beim automatisierten Fahren „mit Überwachung“ werden vom automatisierten Fahrsystem das **Beschleunigen und Bremsen** und das **Lenken** übernommen. Dabei **muss** der Fahrer das System **dauerhaft überwachen** und jederzeit zur vollständigen Übernahme der Fahrzeugführung bereit sein.

Wie zufrieden sind Sie mit der eben verwendeten Automatisierung („mit Überwachung“)?

gar nicht zufrieden ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 sehr zufrieden

Können Sie sich vorstellen, eine Automatisierung („mit Überwachung“) in Zukunft in Ihrem Alltag zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 auf jeden Fall

Was müsste noch verbessert werden, damit Sie das System definitiv nutzen würden?

Bitte beantworten Sie die folgenden Fragen in Bezug auf das automatisierte Fahren („mit Überwachung“) **OHNE die manuelle Übernahme der Fahrzeugsteuerung zu bewerten.**

Geistige Anforderung

Wie hoch waren die geistigen Anforderungen (z.B. Denken, Entscheiden, Erinnern, Hinsehen, Suchen) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Körperliche Anforderung

Wie hoch waren die körperlichen Anforderungen (z.B. Ziehen, Drücken, Drehen, Steuern) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Zeitliche Anforderung

Wie hoch war das Tempo, mit der die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten? War die Aufgabe langsam und geruhsam oder schnell und hektisch?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Aufgabenerfüllung

Wie erfolgreich haben Sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt?

gut ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 schlecht

Anstrengung

Wie sehr mussten Sie sich anstrengen, um Ihre Leistung zu erreichen?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Frustration

Wie verunsichert, entmutigt, gereizt und verärgert fühlten Sie sich während der Aufgabe?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Bitte beantworten Sie die folgenden Fragen **NUR für die manuelle Übernahme der Fahrzeugsteuerung** wenn zuvor automatisiert („mit Überwachung“) gefahren wurde.

Geistige Anforderung

Wie hoch waren die geistigen Anforderungen (z.B. Denken, Entscheiden, Erinnern, Hinsehen, Suchen) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Körperliche Anforderung

Wie hoch waren die körperlichen Anforderungen (z.B. Ziehen, Drücken, Drehen, Steuern) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Zeitliche Anforderung

Wie hoch war das Tempo, mit der die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten? War die Aufgabe langsam und geruhsam oder schnell und hektisch?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Aufgabenerfüllung

Wie erfolgreich haben Sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt?

gut ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 schlecht

Anstrengung

Wie sehr mussten Sie sich anstrengen, um Ihre Leistung zu erreichen?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Frustration

Wie verunsichert, entmutigt, gereizt und verärgert fühlten Sie sich während der Aufgabe?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Sie haben die nachfolgenden Fragen bereits **vor** dem Fahrversuch einmal beantwortet. Bitte beantworten Sie diese nun noch einmal in Bezug auf die **Erfahrungen**, die Sie gerade mit dem System gemacht haben.

Können Sie sich generell vorstellen, in Ihrem Alltag ein solches automatisiertes System zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 auf jeden Fall

Wie ist Ihre Zustimmung zu folgenden Aussagen (**bezogen auf Automatisierung „mit Überwachung**):

Aussage	geringe Zustimmung 1	2	3	4	5	6	starke Zustimmung 7
1. Die Automation ist irreführend.							
2. Die Automation verhält sich auf eine unvorhersehbare Weise.							
3. Ich misstrauere den Aktionen, Absichten oder Konsequenzen der Automation.							
4. Ich bin vorsichtig gegenüber der Automation.							
5. Die Handlungen der Automation führen zu nachteiligen oder schädlichen Konsequenzen.							
6. Ich traue mir zu, die Automation zu nutzen.							
7. Die Automation ist sicher.							
8. Die Automation ist glaubwürdig.							
9. Ich kann mich auf die Automation verlassen.							
10. Die Automation ist zuverlässig.							
11. Ich kann der Automation vertrauen.							
12. Ich bin mit der Automation vertraut.							

Inwieweit treffen die folgenden Aussagen auf Ihre Gefühle während der Aufgabe zu?

Gefühl	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll und ganz zu 5
1. unzufrieden					
2. aufmerksam					
3. deprimiert					
4. traurig					
5. aktiviert					
6. ungeduldig					
7. genervt					
8. ärgerlich					
9. gereizt					
10. missgelaunt					

Inwieweit treffen die folgenden Aussagen auf Ihre Gedanken während der Aufgabe zu?

Gedanken	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll und ganz zu 5
1. Ich möchte die Aufgabe erfolgreich ausführen.					
2. Ich bin motiviert, die Aufgabe auszuführen.					
3. Ich reflektiere mich selbst.					
4. Ich denke über mich selbst nach.					
5. Ich bin mir meiner Fähigkeiten sicher.					

6. Ich bin besorgt, was andere Leute von mir denken.					
7. Ich mache mir Gedanken darüber, was ich für einen Eindruck mache.					
8. Ich erwarte, dass ich die Aufgabe kompetent ausübe.					
9. Insgesamt habe ich das Gefühl, die Kontrolle zu haben.					
10. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben.					
11. Ich habe darüber nachgedacht, wie ich mich fühlen würde, wenn ich eine Rückmeldung bekäme, wie gut ich bei der Aufgabe war.					

Nachfolgend erhalten Sie noch einmal die Automatisierungs-Beschreibung („mit fahrfremden Tätigkeiten“):

Information zu automatisiertem Fahren („mit fahrfremden Tätigkeiten“):

Beim automatisierten Fahren in „mit fahrfremden Tätigkeiten“ werden vom automatisierten Fahrsystem das **Beschleunigen und Bremsen** und das **Lenken** übernommen. Das System muss **nicht dauerhaft überwacht** werden, d.h. der Fahrer darf sich währenddessen mit verschiedenen **anderweitigen (fahrfremden) Tätigkeiten** beschäftigen. Sobald das System an einer **Leistungsgrenze** angelangt ist oder ein **technischer Defekt** auftritt, muss der Fahrer die Fahrzeugführung mit einem Zeitpuffer von wenigen Sekunden übernehmen.

Wie zufrieden sind Sie mit der eben verwendeten Automatisierung („mit fahrfremden Tätigkeiten“)?

gar nicht zufrieden ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 sehr zufrieden

Können Sie sich vorstellen, eine Automatisierung („mit fahrfremden Tätigkeiten“) in Zukunft in Ihrem Alltag zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 auf jeden Fall

Was müsste noch verbessert werden, damit Sie das System definitiv nutzen würden?

Bitte beantworten Sie die folgenden Fragen in Bezug auf das automatisierte Fahren („mit fahrfremden Tätigkeiten“) **OHNE die manuelle Übernahme der Fahrzeugsteuerung zu bewerten.**

Geistige Anforderung

Wie hoch waren die geistigen Anforderungen (z.B. Denken, Entscheiden, Erinnern, Hinsehen, Suchen) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Körperliche Anforderung

Wie hoch waren die körperlichen Anforderungen (z.B. Ziehen, Drücken, Drehen, Steuern) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Zeitliche Anforderung

Wie hoch war das Tempo, mit der die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten? War die Aufgabe langsam und geruhsam oder schnell und hektisch?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Aufgabenerfüllung

Wie erfolgreich haben Sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt?

gut ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 schlecht

Anstrengung

Wie sehr mussten Sie sich anstrengen, um Ihre Leistung zu erreichen?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Frustration

Wie verunsichert, entmutigt, gereizt und verärgert fühlten Sie sich während der Aufgabe?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Bitte beantworten Sie die folgenden Fragen **NUR für die manuelle Übernahme der Fahrzeugsteuerung** wenn zuvor automatisiert („mit fahrfremden Tätigkeiten“) gefahren wurde.

Geistige Anforderung

Wie hoch waren die geistigen Anforderungen (z.B. Denken, Entscheiden, Erinnern, Hinsehen, Suchen) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Körperliche Anforderung

Wie hoch waren die körperlichen Anforderungen (z.B. Ziehen, Drücken, Drehen, Steuern) der Aufgabe?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Zeitliche Anforderung

Wie hoch war das Tempo, mit der die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten? War die Aufgabe langsam und geruhsam oder schnell und hektisch?

sehr niedrig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr hoch

Aufgabenerfüllung

Wie erfolgreich haben Sie die geforderte Aufgabe Ihrer Ansicht nach durchgeführt?

gut ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 schlecht

Anstrengung

Wie sehr mussten Sie sich anstrengen, um Ihre Leistung zu erreichen?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Frustration

Wie verunsichert, entmutigt, gereizt und verärgert fühlten Sie sich während der Aufgabe?

sehr wenig ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 sehr stark

Sie haben die nachfolgenden Fragen bereits **vor** dem Fahrversuch einmal beantwortet. Bitte beantworten Sie diese nun noch einmal in Bezug auf die **Erfahrungen**, die Sie gerade mit dem System gemacht haben.

Können Sie sich generell vorstellen, in Ihrem Alltag ein solches automatisiertes System zu nutzen?

auf keinen Fall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 auf jeden Fall

Wie ist Ihre Zustimmung zu folgenden Aussagen (**bezogen auf Automatisierung „mit fahrfremden Tätigkeiten**):

Aussage	geringe Zustimmung 1	2	3	4	5	6	starke Zustimmung 7
1. Die Automation ist irreführend.							
2. Die Automation verhält sich auf eine unvorhersehbare Weise.							
3. Ich misstrau den Aktionen, Absichten oder Konsequenzen der Automation.							
4. Ich bin vorsichtig gegenüber der Automation.							
5. Die Handlungen der Automation führen zu nachteiligen oder schädlichen Konsequenzen.							
6. Ich traue mir zu, die Automation zu nutzen.							
7. Die Automation ist sicher.							
8. Die Automation ist glaubwürdig.							
9. Ich kann mich auf die Automation verlassen.							
10. Die Automation ist zuverlässig.							
11. Ich kann der Automation vertrauen.							
12. Ich bin mit der Automation vertraut.							

Inwieweit treffen die folgenden Aussagen auf Ihre Gefühle während der Aufgabe zu?

Gefühl	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll und ganz zu 5
1. unzufrieden					
2. aufmerksam					
3. deprimiert					
4. traurig					
5. aktiviert					
6. ungeduldig					
7. genervt					
8. ärgerlich					
9. gereizt					
10. missgelaunt					

Inwieweit treffen die folgenden Aussagen auf Ihre Gedanken während der Aufgabe zu?

Gedanken	Trifft überhaupt nicht zu 1	2	3	4	Trifft voll und ganz zu 5
1. Ich möchte die Aufgabe erfolgreich ausführen.					
2. Ich bin motiviert, die Aufgabe auszuführen.					
3. Ich reflektiere mich selbst.					
4. Ich denke über mich selbst nach.					
5. Ich bin mir meiner Fähigkeiten sicher.					

6. Ich bin besorgt, was andere Leute von mir denken.					
7. Ich mache mir Gedanken darüber, was ich für einen Eindruck mache.					
8. Ich erwarte, dass ich die Aufgabe kompetent ausübe.					
9. Insgesamt habe ich das Gefühl, die Kontrolle zu haben.					
10. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben.					
11. Ich habe darüber nachgedacht, wie ich mich fühlen würde, wenn ich eine Rückmeldung bekäme, wie gut ich bei der Aufgabe war.					

Abschließend bitten wir Sie noch folgende Fragen zu beantworten:


Auf welche Art und Weise wurde Ihrer Einschätzung nach das automatisierte Fahren in diesem Versuch technisch umgesetzt?

Wie viele Stunden haben Sie in der letzten Nacht geschlafen? _____ Stunden



Haben Sie heute Koffein konsumiert? ☐ ja, zuletzt um _____ Uhr
☐ nein


Vielen Dank für Ihre Teilnahme!

Anhang E. Instruktion für die manuelle Fahrt (Baselinefahrt) in Experiment 1 und 2



Runde 1 + 2:  4x Spurwechsel pro Runde

Runde 3 + 4: rechte Spur, immer wiederholen:

von  auf  abbremsen und beschleunigen

Runde 5 + 6:  2x Spurwechsel pro Runde

Runde 7 + 8: mittlere Spur, immer wiederholen:

von  auf  beschleunigen und abbremsen

Runde 9:  konstant, rechte Spur